

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 31, 2002, 18:43:32 ; Search time 1929.4 Seconds
(without alignments)
3191.319 Million cell updates/sec

Title: US-09-641-831-1
Perfect score: 573
Sequence: 1 atgagcaggaccactgaaga.....gagcctattaagatgata 573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	253	44.2	605	10 AW369663	AW369663 QV0-BT004
C 2	246.6	43.0	437	11 BF591083	BF591083 7h52f04.x
C 3	238	41.5	359	10 AW294613	AW294613 UI-H-B12-
C 4	204	35.6	300	10 AW590275	AW590275 hg42f03.x
C 5	158.4	27.6	265	10 A1300504	A1300504 qo19c11.x
C 6	150	26.2	649	11 BG563384	BG563384 602582446
7	100.8	17.6	1121	12 AK018481	AK018481 Mus muscu
8	91.2	15.9	321	10 BS504666	BS504666 BB504666
9	84.8	14.8	484	10 AU170966	AU170966 AU170966
10	84.6	14.8	481	11 BF463332	BF463332 UI-W-C30P
11	82.6	14.4	231	10 AV368091	AV368091 AV368091
12	75	13.1	597	11 BF057621	BF057621 7k46h06.x

13	70.8	12.4	558	11 BF434691	BF434691 7p03b0..
14	66.6	11.6	354	10 AW001038	AW001038 wr91h03..
15	64.6	11.3	383	13 AQ244687	AQ244687 HS_2045.A
16	53.6	9.1	932	13 CNS03SP4	AL258817 Tetraodon
17	52.4	9.1	614	13 AZ761049	AZ761049 1M0555J09
18	44.2	7.7	866	13 CNS05R29	AL350010 Tetraodon
19	43.6	7.6	1092	13 CNS0372N	AL231980 Tetraodon
20	40.6	7.1	379	10 AW356254	AW356254 38503 MAR
21	39.8	6.9	378	11 BF003124	BF003124 7g55e02.x
22	39.4	6.8	586	13 AQ697936	AW356254 38503 MAR
23	39	6.8	412	11 W24504	AQ697936 HS_5536.B
24	38.4	6.7	334	10 AA259846	W24504 zb61a09.r1
25	38.4	6.7	553	10 AW845396	AA259846 va36g01.r
26	38.4	6.7	1553	12 AK016447	AW845396 CM4-CT004
27	38.2	6.7	458	10 A1825748	AK016447 Mus muscu
28	38.2	6.7	882	11 BF693457	A1825748 wb69b08.x
29	37.8	6.6	387	10 AW058549	BF693457 602082111
30	37.4	6.5	290	10 AU168952	AW058549 wx23c10.x
31	37.4	6.5	437	10 A1309314	AU168952 AU168952
32	37.4	6.5	465	10 A1309314	A1309314 qo80b08.x
33	37.2	6.5	398	10 AA111914	AA111914 zh53c03.f
34	37.2	6.5	473	10 AW083037	AW083037 xc05f08.x
35	37.2	6.5	924	10 BE730550	BE730550 601562213
36	37.2	6.5	946	13 CNS02DER	AL192348 Tetraodon
37	37	6.5	521	11 BF016403	BF016403 uv40g09.y
38	36.8	6.4	185	10 AW230070	AW230070 up25b06.y
39	36.8	6.4	288	11 W12275	W12275 ma06b09.r1
40	36.8	6.4	371	11 W20834	W20834 mb93a02.r1
41	36.8	6.4	479	10 AA855804	AA855804 vw75a10.f
42	36.8	6.4	499	10 AA764505	AA764505 vp05a05.f
43	36.8	6.4	515	10 AW743037	AW743037 up62a05.y
44	36.8	6.4	532	10 A1509426	A1509426 va36g01.y
45	36.8	6.4	558	10 AA869064	AA869064 vq30c11.f

ALIGNMENTS

RESULT 1
AW369663/c 605 bp mRNA EST 04-FEB-2000
LOCUS QV0-BT0041-011199-039-ell BT0041 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW369663
ACCESSION AW369663
VERSION AW369663.1 GI:6874317
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 605)
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-BT0041-011199-039-ell&t3=1999-11-01&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 586.
Location/Qualifiers
1. 605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0041"

FEATURES
source

```

/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      163 a      132 c      144 g      166 t
ORIGIN

```

```

Query Match      44.2%; Score 253; DB 10; Length 605;
Best Local Similarity 100.0%; Pred. No. 3.7e-62;
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 cagcctctccccacgggtcaactgttagctgagttgacctcaacatccccagagat 380
|||||
DB 455 CAGCCTCTCCACCACGGTCACCTGTTAGCTCGAGTTTGACCCCTCAACATCCCCAGGAT 396
|||||
QY 381 gctcttaggactcgaacctcttagtctctactctgtagcagccatgagtggag 440
|||||
DB 395 GCCTCTAGACTCGAACCTCAATTAGTGTCTACTGCTGATGACGCCATGAGTGGAG 336
|||||
QY 441 ctgtcaccaccttgggtgaagtgtgtctccacgtctccacctgggcagcacgatgcc 500
|||||
DB 335 CTGTGACCACTTGGGTGAAGTGTGTCTCCAGTCTCCAGTGGGAGCAGCATGGCC 276
|||||
QY 501 acagaaagttgtattatccccaccacagcacacacatgcagaaatttcaaaagccta 560
|||||
DB 275 ACAGAAAGTGTGTATCTACCCACCAGCAGCAACACATGCAGAAATTTCAAAAGAGCCTA 216
|||||
QY 561 tttaagtagatga 573
|||||
DB 215 TTTAAGTAGATGA 203
|||||

```

```

RESULT 2
BF591083/c
LOCUS      7h52f04.x1 NCI_CGAP_Col16 Homo sapiens cDNA clone IMAGE:3319615 3',
DEFINITION mRNA sequence.
ACCESSION BF591083
VERSION    BF591083.1 GI:11683407
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 437)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps@mail.nih.gov
            Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
            , Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            DNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40up from Gibco
            High quality sequence stop: 424.
            Location/Qualifiers
            1. 437
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3319615"
            /clone_lib="NCI_CGAP_Col16"
            /tissue_type="colon tumor, RER+"

```

```

FEATURES
source
1. 437
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3319615"
/clone_lib="NCI_CGAP_Col16"
/tissue_type="colon tumor, RER+"

```

```

/lab_host="DH10B"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Col10 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo.
BASE COUNT      117 a      100 c      101 g      119 t
ORIGIN

```

```

Query Match      43.0%; Score 246.6; DB 11; Length 437;
Best Local Similarity 98.4%; Pred. No. 2.3e-60;
Matches 249; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 321 cagcctctccccacgggtcaactgttagctgagttgacctcaacatccccagagat 380
|||||
DB 366 CAGCCTCTCCACCACGGTCACCTGTTAGCTCGAGTTTGACCCCTCAACATCCCCAGGAT 307
|||||
QY 381 gctcttaggactcgaacctcttagtctctactctgtagcagccatgagtggag 440
|||||
DB 306 GCCTCTAGACTCGAACCTCAATTAGTGTCTACTGCTGATGACGCCATGAGTGGAG 247
|||||
QY 441 ctgtcaccaccttgggtgaagtgtgtctccacgtctccacctgggcagcacgatgcc 500
|||||
DB 246 CTGTGACCACTTGGGTGAAGTGTGTCTCCAGTCTCCAGTGGGAGCAGCATGGCC 187
|||||
QY 501 acagaaagttgtattatccccaccacagcacacacatgcagaaatttcaaaagccta 560
|||||
DB 186 ACAGAAAGTGTGTATTTACCCACCAGCAGCAACACATGCAGAAATTTCAAAAGAGCCTA 127
|||||
QY 561 tttaagtagatga 573
|||||
DB 126 TTTAAGTAGATGA 114
|||||

```

```

RESULT 3
AW294613/c
LOCUS      AW294613      359 bp      mRNA      EST      16-JAN-2000
DEFINITION UI-H-B12-ahk-h-11-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
IMAGE:2727404 3', mRNA sequence.
ACCESSION AW294613
VERSION    AW294613.1 GI:6701249
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 359)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 forward
            POLYA=yes.
            Location/Qualifiers
            1. 359
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2727404"
            /clone_lib="NCI_CGAP_Sub4"

```

```

FEATURES
source
1. 359
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2727404"
/clone_lib="NCI_CGAP_Sub4"

```



```

/clone_11b-"RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGGAGAGAGAGCGCCGCACTCGAGTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using reverse transcriptase thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adaptor of sequence [5',
GAGGAGAGAGATTCTCGATTATTAATTAATATCCCCCCCCCC 3'}. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pbluescript KS(+) after bulk excision from Lambda PLoC I."

```

Query Match	15.9%	Score 91.2	DB 10	Length 321
Best Local Similarity	71.4%	Pred. No. 1.1e-15		
Matches 120	Conservative 0	Mismatches 48	Indels 0	Gaps 0

0y 3 gaagaggacacactaaagcttcacaagctcagtagcgccacattaaactctaaacacgqgcac 62
Db 76 gatgaagattccggagaaacttggagagaaacctgaactccaccccccaaccctcagtagagacc 135
0y 63 caaggaagctacattactctgagagcattccctcggagagagagctccctcgaggtttac 122
136 caccgagagattcgtctactctggaaagccctcttgaagagaaagggctctcccgggcttcac 195b

Oy 123 tctaaaggttgccctgagcacygaacatcatcatctaagt 170
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 CCTGAAGGTGGCCCTGGAGCGTGGAGAACCCTAATCATTTCCCAAGT 243

RESULT 9
AV170966

LOCUS	484 bp	mRNA	EST	29-JAN-2001
DEFINITION	AU170966	cdna	Oryzias latipes	cdna clone br7042, mRNA sequence.

VERSION	AUI70966.1	GI:12593037
KEYWORDS	EST.	
SOURCE	Japanese medaka.	

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Oryziatidae; Oryziatidae.

AUTHORS Mita,K., Ishikawa,Y. and Yanauchi,M.
TITLE Establishment of cDNA database of medaka *Oryzias latipes*
JOURNAL Unpublished (2001)
COMMENT Contact: Mita K

National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 265-8555, Japan
Email: knita@nirs.go.jp
method: uni-directional sequence direction: sequenced from T3 primer
(5' -> 3').

FEATURES	source	location/Qualifiers
1. . 484		
/organism="Oryzias latipes		
/string="HN1"		
/db_xref="taxon:8090"		
/clone="br7042"		
/clone_lib="O1-br-ad cdna"		

```

/sex="female/male mixed"
/tissue_type="brain"
/dev_stage="adult"
130 a 114 c 150 g 90 t
BASE COUNT
ORIGIN

```

Query Match	14.8%	Score 84.8	DB 10	Length 484
Best Local Similarity	65.9%	Pred. No. 8.5e-14		
Matches 141, Conservative	0	Mismatches 67	Indels 6	Gaps 1

QY 100 ggaagagctccctgagggttactctaagaagtgcctgagacgagaaaccatc 15
|| || || || || || || || || || || || || || || || || || || ||
Db 277 GGGGCGCCCATGGGGTTCACCTCAGAGAGAAGACTGGAGCATCAGAGACCCTGTGC 33

OY 160 atcttaagtcgaagaaggggccaacgacacacctgagtcccaactgcagcgtggg 21

D8 || | | | | | | | | | | | | | | | | |
 || | | | | | | | | | | | | | | | | |
Db 337 ATAAACCAGGTGGAGAAGGCGCAGCAAGCCGCAGGCCGTAGT-----CTTCAGTGGA 39
 |||

[illegible]

QY 280 ctgctgaaggtatcctacaagaccctagctgg 313
||||||| | ||| ||||| ||| | |||
|||

RESULT 10

[illegible]

ACCESSION	BF463332
VERSION	BF463332.1
KEYWORDS	GI:11532515
	EST.

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 461)	Bonaldo, M.F., Lennon, G. and Soares, M.B.	Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL
MEDLINE
COMMENT
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Chin, H
National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706

Fax: 001 443 3630
Email: MEST@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.

non-redundant arrays of BMAP cDNAs whose availability will be

considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA-No.

FEATURES	SOURCE	Location/Qualifiers
	1..481	
		/organism="Mus musculus"
		/strain="C57BL/6J"
		/db_xref="taxon:10090"
		/clone="U1-M-CG0p-hm-c-05-0-01"
		/clone_lib="NIH BMAP Ref4.S2"
		/lab_host="DH10B (Life Technologies)"
		/note="vector: pUT3d-Pac (Pharmacia) with a modified
		polylinker; Site_1: Not I; Site_2: Eco RI; The
		NIH BMAP Ref4.S2 library is a subcloned library,

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome
Clone distribution: NCI-CGAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 330
Std Error: 0.00
Seq primer: -40up from Gibco.

FEATURES	Location/Qualifiers
source	1. .354

```
1. 354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2495093"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="PH108"
```

BASE COUNT	82 a	85 c	80 g	107 t
ORIGIN	falling bombards.			

	Query Match	11.68;	Score 66.6;	DB 10;	Length 354;
	Best Local Similarity	60.8;	Prism. No. 1.3e-08;		
	Matches 144;	Conservative	0;	Mismatches 89;	Indels 4; Gaps 2;
QY	111	ctgggggtttactctaaagggtggccctgggagcagagaaaccattaatca-tctctaaagg	169		
Db	354	CTGGGGCTTCACCTTAAGGGGCTCGGAACCTGTGAACGGCTCCCAAGTCTTAAGA	295		
QY	170	tgaagaagggggcaaacagacacacctgagctcccaactgcagcgtctggggatgaggttg	229		
Db	294	TTGAAGATGGGGCAAGGCAG---CTTGTGCCAAGAGATGAGGACTGGTGATGAGCTGG	238		
QY	230	tgcacatcatcagtgactctgagcagctcagaaaggaggcagtttccctcggtgaaag	289		
Db	237	TGAATATCAATGGCACTCCATTATATGCTCCGCCAAGAGGCCCTCATCTCATCAAG	178		
QY	290	gattctacaagaccctcagcgttgtagtcgcagcctctcccacacccggtcaactgta	346		
Db	177	GCCTCTCCGGATTCTCAAGCTGATTGTTCAGAGACTAAGTTTGCTGGCCCCGTGTA	121		

RESULT	15
AQ244687	
LOCUS	383 bp DNA GSS 03-OCT-1998
DEFINITION	HS_2045_A2_H11_T7 Cit Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2045 Col=22 Row=O, DNA sequence.
ACCESSION	AQ244687
VERSION	AQ244687.1
KEYWORDS	GI:3691261
SOURCE	GSS.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 383) Mahaliras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keiler,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahaliras GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 31, 2002, 19:16:09 ; Search time 1929.4 Seconds
(without alignments)
1821.224 Million cell updates/sec

Title: us-09-641-831-3
Perfect score: 327
Sequence: 1 atgatgagaccactgaaga.....ggctgtagtgcagattga 327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 11351937 seqs, 5372889281 residues 22703874

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl1:*
11: gb_estl2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148.4	45.4	649	11	BG563384 602582446
2	100.8	30.8	1121	12	AK018481 Mus muscu
3	91.2	27.9	321	10	BB504666 BB504666
4	84.8	25.9	481	10	AU170966 AU170966
5	84.6	25.9	481	11	BF463332 UI-M-CGOp
6	82.6	25.3	231	10	AV368091 AV368091
7	77.5	22.9	597	11	BF057621 7k46h06.x
8	70.8	21.7	558	11	BF434691 7p03b06.x
9	65	19.9	383	13	AQ244687 HS_2045_A
10	62.8	19.2	354	10	AW001038 wr9lh03.x
11	54	16.5	932	13	CNS03SP4
12	45	13.8	866	13	CNS050R29

13	44.4	13.6	1092	13	CNS0372N
14	42.2	12.9	458	10	AI825748
15	40.6	12.4	379	10	AQ356254
16	39.4	12.0	586	13	AO697936
17	39	11.9	412	11	W24504
18	38.4	11.7	334	10	AA259846
19	38.4	11.7	553	10	AA845396
20	38.4	11.7	1553	12	AK016447
21	38.2	11.7	882	11	BF693457
22	37.4	11.4	290	10	AU168952
23	37.2	11.4	398	10	AA111914
24	37	11.3	378	11	BF003124
25	37	11.3	521	11	BF016403
26	36.8	11.3	185	10	AW230070
27	36.8	11.3	288	11	W12275
28	36.8	11.3	371	11	W20834
29	36.8	11.3	479	10	AA855804
30	36.8	11.3	499	10	AA764505
31	36.8	11.3	515	10	AW743037
32	36.8	11.3	532	10	AI509426
33	36.8	11.3	558	10	AA869084
34	36.8	11.3	581	11	BG922902
35	36.6	11.2	212	10	AL603379
36	36.6	11.2	300	10	AU098787
37	36.6	11.2	309	11	D31139
38	36.6	11.2	321	10	AA374357
39	36.6	11.2	323	11	BG407069
40	36.6	11.2	375	10	AA112776
41	36.6	11.2	415	10	AA329300
42	36.6	11.2	455	10	AA446635
43	36.6	11.2	456	11	H72671
44	36.6	11.2	463	11	BF001694
45	36.6	11.2	467	11	W07582

ALIGNMENTS

RESULT 1
BG563384 649 bp mRNA 10-APR-2001
LOCUS 602582446F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710299 5',
DEFINITION mRNA sequence.
ACCESSION BG563384
VERSION BG563384.1 GI:13571036
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 649)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC41548 row: 1 column: 12
High quality sequence stop: 548.
Location/Qualifiers
1. 649
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4710299"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1."

BB504666					
LOCUS 321 bp mRNA EST 26-JUL-2000					
DEFINITION RIKEN full-length enriched, 0 day neonate kidney Mus musculus cDNA clone D830049123 3' similar to AF194421 Mus musculus PDZ domain actin binding protein Shroom mRNA, mRNA sequence.					
ACCESSION BB504666.1 GI:9493460					
VERSION BB504666.1					
KEYWORDS house mouse.					
ORGANISM Mus musculus					
REFERENCE 1 (bases 1 to 321)					
AUTHORS Konno H., Aizawa, K., Akahira, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, F., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kal, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Mikai, H., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sugabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.					
RIKEN Mouse ESTs (Konno, H., et al.)					
Unpublished (2000)					
Contact: Yoshihide Hayashizaki					
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute					
The Institute of Physical and Chemical Research (RIKEN)					
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan					
Tel: 81-45-503-9222					
Fax: 81-45-503-9216					
Email: genome-res@gsr.riken.go.jp,					
URL: http://genome.gsc.riken.go.jp/					
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.					
Thermotranscription and thermotranscription of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)					
Itoh, M., Kitsumaru, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.					
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)					
Carninci, P. and Hayashizaki, Y.					
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)					
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.					
FEATURES					
Source Location/Qualifiers					
1..321					
/organism="Mus musculus"					
/db_xref="taxon:10090"					
/clone_lib="RIKEN full-length enriched, 0 day neonate kidney"					
/tissue_type="kidney"					
/dev_stage="0 day neonate"					
/lab_host="DH10B"					
/note="Site 1: Salt; Site 2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'					
GAGAGAGCGGGCCGACTCTTGTATTTTTCCTTAAGTGCCTGGAGCAACCATTAACTAATCCCCCCC 3'] cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'					
GAGAGAGATTTCGAGTGAATAATTAATCCGCCCCC 3'] cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1."					
BASE COUNT 64 a 89 c 85 t					
ORIGIN					
Query Match 27.9%; Score 91.2; DB 10; Length 321;					
Best Local Similarity 71.4%; Pred. No. 3.9e-15;					
Matches 120; Conservative 0; Mismatches 48; Indels 0; Gaps 0;					
QY 3 gatgaggaccactgaagactctccacaacgttagtgccacattaaacttaaacgcggcacc 62					
Db 76 GATGAGATTCCGGAGAATCTGGAGGAGCCTAGCTCCACCCCAACCCAGTAGGACCCC 135					
QY 63 caagggaagtcatattatctctggagcgattcctggaggaggagagctcctgggtttac 122					
Db 136 CACGGAGAGATTTCGTTTCATCTCTGGAAGCGCTCTAGAGGAGGGGCTCCCGGGGCTTCAC 195					
QY 123 ttctaaagggtgacctgagcagcgagcaaccattaatctctctaaggt 170					
Db 196 CCTGAAGGTGGCTTGAGCGCTGGAGAACCGTTATCATTTCCAGGT 243					
RESULT 4					
LOCUS AU170966 484 bp mRNA EST 29-JAN-2001					
DEFINITION AU170966 Ol-br-ad cDNA Oryzias latipes cDNA clone br7042, mRNA sequence.					
ACCESSION AU170966					
VERSION AU170966.1 GI:12593037					
KEYWORDS EST.					
SOURCE Japanese medaka.					
ORGANISM Oryzias latipes					
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acantopecterygii; Percormorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.					
1 (bases 1 to 484)					
Mita, K., Ishikawa, Y. and Yamauchi, M.					
Establishment of cDNA database of medaka, Oryzias latipes					
Unpublished (2001)					
Contact: Mita K					
Genome Research Group					
National Institute of Radiological Sciences					
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan					
Email: knita@nirs.go.jp					
Method: uni-directional sequence direction:sequenced from T3 primer (5' -> 3').					
FEATURES					
Location/Qualifiers					
1..484					
/organism="Oryzias latipes"					
/strain="HNI"					
/db_xref="taxon:8090"					
/clone_lib="br7042"					
/clone_lib="Ol-br-ad cDNA"					
/sex="female/male mixed"					
/tissue_type="brain"					
/dev_stage="adult"					
BASE COUNT 130 a 114 c 150 g 90 t					
ORIGIN					
Query Match 25.9%; Score 84.8; DB 10; Length 484;					
Best Local Similarity 65.9%; Pred. No. 2.6e-13;					
Matches 141; Conservative 0; Mismatches 67; Indels 6; Gaps 1					
QY 100 gagagagctccctgggttttactctaaagggtgctcgagcgaggaaccattatc 159					
Db 277 GGGGGGCCCGCCATGGGGGTTACCTCTAGAGGAGCTGGAGCATCAGGACCGCTTGTC 336					
QY 160 atctctaaggttcqaaagggggcaaagcacacctgagctccaactcagctggg 219					

Db 337 ATACCAAGGTGAGGAGGACCAAGCGCAGCCGTGAGT-----CTTCAAGTGGGA 390
 Oy 220 gatgaggtgtgtgacatcaatgaggtgactctgagcagctccagaagagcagttcc 279
 Db 391 GAGGAGCTCTCAACATTAACCCAGATTCCCTGAGGCGGTACCGAGAGGCCATCTCG 450
 Oy 280 ctgtgtaaaagatcctacaagacctcaggcttg 313
 Db 451 CTGTGAGAGGCTCCACCAAGACCTGACTCTG 484

RESULT 5
 BF463332 481 bp mRNA EST 04-DEC-2000
 LOCUS UI-M-CGDP-dnm-c-05-0-UI.s1 NIH_BMAP_ret4-S2 Mus musculus cDNA clone
 DEFINITION UI-M-CGDP-dnm-c-05-0-UI 3', mRNA sequence.
 ACCESSION BF463332
 VERSION BF463332.1 GI:11532515
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 481)
 AUTHORS Donald, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: MEST@mail.nih.gov

Oligo-dT track not found. Not a site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
 clones from RESEARCH GENETICS. It should be noted that Bento Soares
 is generating a small number of additional specialized
 non-redundant arrays of BMAP cDNAs whose availability will be
 considered under appropriate and limited collaborative arrangements
 Seq primer: M13 forward
 POLYA-No.

FEATURES

source
 1. 481
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-CGDP-dnm-c-05-0-UI"
 /clone_1lb="NIH_BMAP_Ret4_S2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NIH_BMAP_Ret4_S2 library is a subtracted library,
 ultimately derived from mouse retina tissue libraries at
 various stages of development. For a detailed description
 of the library from which this clone was derived, please
 visit our web site at brainest.eng.uiowa.edu.
 TAG_SEQ=None found"
 BASE COUNT 125 a 140 c 115 g 101 t
 ORIGIN

Query Match 25.9%; Score 84.6; DB 11; Length 481;
 Best Local Similarity 69.5%; Pred. No. 2.9e-13;
 Matches 130; Conservative 0; Mismatches 54; Indels 3; Gaps 1;
 Oy 131 gtggcctgagcaggaacattcaatcctcaggtcgaagaggggagcaagag 190
 Db 1 GCGGCGCGAGCAGTGGCAGCTCTGTCATCCACCAAGATTGAGAGGAGCAAGCTG 60

Oy 191 aacccctgagctccaaactcgaagctgggagtgatgtgtgtcacatcaatgaggtgactc 250
 Db 61 CAGCAGTGGAC---AAGTACTTGTCTGGAGATGAGATGAGTGGCATTAATGATGTGAGTC 117
 Oy 251 tgaagagctccagaagagagcagttccctggtgaagagatcctacaagacctcaggc 310
 Db 118 TCTCAGGGTTTCAGACAAAGACGCAATTTCCTGCTGTAAGGCTCCACCAAGACCTGAAGC 177
 Oy 311 tggtagt 317
 Db 178 TTGTGTT 184

RESULT 6
 AV368091 231 bp mRNA EST 14-NOV-1999
 LOCUS AV368091 RIKEN full-length enriched, 16 days embryo lung Mus
 DEFINITION Musculus cDNA clone 8430440M04 3', mRNA sequence.
 ACCESSION AV368091
 VERSION AV368091.1 GI:6415738
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 231)
 AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
 Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
 Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,
 C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Owa, C., Ozawa, Y., Saito, H., Sano, K., Sato, K., Shibata, K., Shibata,
 Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H.,
 Suzuki, H., Takahashi, F., Tateo, M., Tomimaga, N., Tsunoda, Y.,
 Watahiki, A., Watanabe, S., Yamamuta, T., Yasunishi, A., Yokota, T.,
 Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 TITLE RIKEN Mouse ESTs (Kono, H., et al. 1999)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp.
 URL: http://genome.gsc.riken.go.jp/
 Sasaki, N., Izawa, M., Watanabe, S., Okazaki, Y. and Hayashizaki,
 Y.
 Matsuyama, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki,
 Y.

Transcriptional sequencing: A method for DNA sequencing using RNA
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

FEATURES

source
 1. 231
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="8430440M04"
 /clone_1lb="RIKEN full-length enriched, 16 days embryo
 lung"
 /sex="mixed"
 /tissue_type="lung"
 /dev_stage="16 days embryo"

```
/lab_host="DH10B"
/notes="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAAAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot - 10.0 and subtraction to Rot - 195.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTAAATTAATCCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI"
BASE COUNT      52 a      56 c      52 g      71 t
ORIGIN
Query Match      25.3%; Score 82.6; DB 10; Length 231;
Best Local Similarity 71.2%; Pred. No. 8.8e-13;
Matches 109; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 18 agacttcacaaagccctagtgccattaaactctaacagggccacaaaggaaggtacat 77
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 AAAATTGTAGCAGCTAGCCCTACTCAAAACCCCTACTAGCAGCCACCCAGAGATTCCT 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 78 ttatctggaggaattctggagagagagctccctgggttttactctaaagggtggcct 137
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TTATATGGAACCCCTCTCCTAGAGGAGGGGATCCCTGGGCCCTTACCCTGAAGGTGGCCT 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 138 gqagcagggagacattatcatctctaaagt 170
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 GGAGCGTGGAGACCGTATTTCATTTCCAAAGGT 153
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 7
BF057621 597 bp mRNA EST 16-OCT-2000
LOCUS 7k46n06.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3478523 3'
DEFINITION similar to SW:APXL_HUMAN Q13796 APICAL-LIKE PROTEIN ; contains
MER22_b2 TAR1 repetitive element ;, mRNA sequence.
ACCESSION BF057621
VERSION BF057621.1 GI:10811517
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
infoimage.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 511.
Location/Qualifiers
1. .597
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone="IMAGE:3478523"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrothecoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAAGTGGAGCGCGCGGACATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      125 a      205 c      158 g      108 t
ORIGIN
Query Match      22.9%; Score 75; DB 11; Length 597;
Best Local Similarity 66.3%; Pred. No. 1.4e-10;
Matches 124; Conservative 0; Mismatches 60; Indels 3; Gaps 1;
QY 131 gtgacctgagcagcagagaccattatcatctctaaagtcgaaggggcaagcag 190
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GCGGCGCGAGCAGCGCGCGCGCTGTCTCATCCACAGATTGAAGAGGCGCAGTAAGC-- 58
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 191 acacctgagctccaaactgcagctgaggtgaggtgtgtgcacatcaatgagtgactc 250
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 -CGCGCGGTGCAGCAAGTACTTGGCTGGAGATGAGATCGCGCATCAATGACATTGGTC 117
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 251 tgagcagctccagaaaggagcagttccctgggtgaagagatcctcacaagacctcaggc 310
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 118 TCTCAGGTTTAGACAGGAGGAGGATTCCTCGTGGAAGGGTCCCTAAGACCCCTGAGC 177
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 tggtagt 317
| | | | |
Db 178 TGGTGT 184
| | | | |
RESULT 8
BF434691 558 bp mRNA EST 29-NOV-2000
LOCUS 7p03b06.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644507 3'
DEFINITION similar to SW:APXL_HUMAN Q13796 APICAL-LIKE PROTEIN ;, mRNA
sequence.
ACCESSION BF434691
VERSION BF434691.1 GI:11446979
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 558)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
infoimage.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 488.
Location/Qualifiers
1. .558
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source
```

```
/clone="IMAGE:3644507"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH108 (phage-resistant)"
/notes="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACCACTGAAGTGGAGCGCGGCACATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      118 a 191 c 147 g 102 t
ORIGIN

Query Match      21.7%; Score 70.8; DB 11; Length 558;
Best Local Similarity 66.3%; Pred. No. 2e-09;
Matches 118; Conservative 0; Mismatches 57; Indels 3; Gaps 1;

QY 140 agcagggagaccattatcatctctgaagtcgaagggggaagcagacacacctga 199
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 AGCAGCGCGAGCGCGTGGTCATCCACCAAGATTGAAGGGCGCATAAGC---CGCGGCGG 57

QY 200 gctccaaactcagcgtgggagtgaggtgtgcacatcaatcgaatgagtcgagcgt 259
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 TCACAAAGTTACTGGCTGAGATGAGATCGTCGGCATCAATGATGCTCTCAGGGT 117

QY 260 ccagaagagcagcttccctggtgaagagtcctcacaagaccctcagcgtgtagt 317
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 TTAGACAGGAAGCGATTTCCTGGTGAAGGGTCCATAGACCCCTGAAGCTGGTGT 175

RESULT 9
LOCUS      AQ244687
DEFINITION HS_2045_A2_H11_T7 C17 Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2045 Col=22 Row=O, DNA sequence.
ACCESSION  AQ244687
VERSION     AQ244687.1 GI:3691281
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Sequence Tagged Connector
Plate: 2045 Row: O Column: 22
Class: BAC ends
High quality sequence stop: 383.
Location/Qualifiers
1..383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=2045 Col=22 Row=O"
/clone_lib="C17 Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in
```

```
BASE COUNT      86 a 92 c 115 g 90 t
ORIGIN

Query Match      19.9%; Score 65; DB 13; Length 383;
Best Local Similarity 64.1%; Pred. No. 7.5e-08;
Matches 98; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 175 gaaggggggaaagcagacacacctgagctccaaactcagctcagctgggaggtgtgcac 234
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 GAAGAGGGCAGTAAGCCGCGGTCGACAAAGTTACTTGGCTGGAGATGAGATCGTCGGC 275

QY 235 atcaatgagtgactctgagcagctccagaagagcagcttccctcgtggaagagatcc 294
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 ATCAATGACATTGCTCTCAGGGTTTAGACAGGAAGCGATTTCCTGGTGAAGGGGTCC 335

QY 295 tacaagacctcagctggtgtagtcgagctga 327
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 CATAAGACCTGAAGCTGGTCTGCTCAAAAGGTAA 368

RESULT 10
LOCUS      AW001038/c
DEFINITION wr1h03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2495093 3'
similar to SW:APXL_HUMAN Q13796 APICAL-LIKE PROTEIN ;, mRNA
sequence.
ACCESSION  AW001038
VERSION     AW001038.1 GI:5847954
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 354)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrrp/image/image.html
Insert Length: 330 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2495093"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH108"
/notes="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT      82 a 85 c 80 g 107 t
ORIGIN
```

Query Match	19.2%;	Score 62.8;	DB 10;	Length 354;	
Best Local Similarity	62.11%;	Pred. No. 3e-07;			
Mismatches	133;	Conservative	0;	Mismatches	77;
				Indels	4;
				Gaps	2;
QY	111	ctgggttttactctaaagggtgctgtgagcagcagagaaaccatttaaca-tctctaagg	169		
Db	354	CTGGGCTTCACCTTAAGGGGGTCTGGAACCTGTGAACCGGCTCCGAGTGTCTAGA	295		
QY	170	tcgaagaagggggaagcagacacccctgagctccaaactcagggctgggaggttg	229		
Db	294	TTGAAGATGAGCAGCAGGAG---CTTTGTCCAGAAGATGAGGACTGGTGATGAGCTGG	238		
QY	230	tgacatcaatgaggtgactctgagcagctcagaaaggagcagtttccctggggaag	289		
Db	237	TGAATATCAATGGCACTCCATTATATGCTCCGCCAGAGGCCCTCATCTTCATCAAG	178		
QY	290	gatcctacaagaccctcaggtgtagtcgcag	323		
Db	177	GCTCCTCCGGATTCTCAAGCTGATTGTTCAGGAG	144		
RESULT 11					
CNS03SP4					
LOCUS		932 bp	DNA	GSS	17-MAY-2000
DEFINITION		Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 054119 of library G from tetraodon nigroviridis, genomic survey sequence.			
ACCESSION	AL258817				
VERSION	AL258817.1	GI:7979829			
KEYWORDS	GSS; genome survey sequence.				
SOURCE	Tetraodon nigroviridis.				
ORGANISM	Tetraodon nigroviridis				
REFERENCE	1 (bases 1 to 932)				
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.				
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 932)				
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.				
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 932)				
AUTHORS	Direct Submission				
TITLE	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases				
JOURNAL	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at				
COMMENT	http://www.genoscope.cns.fr/Tetraodon.				
FEATURES	Location/Qualifiers				
source	1. .932				
	/organism="Tetraodon nigroviridis"				
	/db_xref="taxon:99883"				
	/clone="054119"				
	/clone_lib="G"				
	/note="Genoscope sequence ID : C08G054AE10SP1-end : PUC-ori"				
BASE COUNT	177 a	260 c	303 g	189 t	3 others
ORIGIN					
Query Match	16.5%;	Score 54;	DB 13;	Length 932;	
Best Local Similarity	66.11%;	Pred. No. 0.0001;			
Mismatches	78;	Conservative	0;	Mismatches	40;
				Indels	0;
				Gaps	0;

QY	208	ctcagagctgggagtgaggttgcacatcaatgaggtgactctgagcagctccgaaag	267		
Db	43	CTCGAGCCGGTGATGAAATGGTGAGCGTGACCGCTCGTCTCAGCGCTCCCGTCAG	102		
QY	268	gagcaggttccctgtgaaagatctctacaagacctcagctgagtagtcgcagatt	325		
Db	103	GAGCCATCTCTGTGTGAAGACTCGCAGAGCTTTGACCTTGGTGTGTCGCAGGT	160		
RESULT 12					
CNS05R29/c					
LOCUS		866 bp	DNA	GSS	26-MAY-2000
DEFINITION		Tetraodon nigroviridis genome survey sequence T7 end of clone 014K24 of library C from tetraodon nigroviridis, genomic survey sequence.			
ACCESSION	AL350010				
VERSION	AL350010.1	GI:8243780			
KEYWORDS	GSS; genome survey sequence.				
SOURCE	Tetraodon nigroviridis.				
ORGANISM	Tetraodon nigroviridis				
REFERENCE	1 (bases 1 to 866)				
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.				
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 866)				
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.				
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 866)				
AUTHORS	Direct Submission				
TITLE	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases				
JOURNAL	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at				
COMMENT	http://www.genoscope.cns.fr/Tetraodon.				
FEATURES	Location/Qualifiers				
source	1. .866				
	/organism="Tetraodon nigroviridis"				
	/db_xref="taxon:99883"				
	/clone="014K24"				
	/clone_lib="C"				
	/note="Genoscope sequence ID : C0TC014BF12C2-end : T7"				
BASE COUNT	181 a	199 c	220 g	210 t	56 others
ORIGIN					

RESULT	13	GENS	15-MAY-2000
LOCUS	CNS0372N	1092 bp	DNA
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-Or1 end of clone 004C04 of library G from Tetraodon nigroviridis, genomic survey sequence.		
ACCESSION	AL231980.1	GI:7890975	
VERSION	AL231980.1	GI:7890975	
KEYWORDS	GSS: genome survey sequence.		
SOURCE	Tetraodon nigroviridis		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodon.		
REFERENCE	1 (bases 1 to 1092)		
AUTHORS	Roest-Crollius,H., Jalllon,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.		
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1092)		
AUTHORS	Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.		
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 1092)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases		
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .		
FEATURES	Location/Qualifiers		
source	1..1092		
	/organism="Tetraodon nigroviridis"		
	/db_xref="taxon:99883"		
	/clone="004C04"		
	/clone_lib="G"		
	/note="Genoscope sequence ID : C093004BB02XEL-end : PUC-Or1"		
BASE COUNT	262 a	294 c	304 g 231 t 1 others
ORIGIN			
Query Match	13.6%	Score 44.4	DB 13; Length 1092;
Best Local Similarity	58.2%	Pred. No. 0.043	Indels 0; Gaps 0
Matches 78; Conservative	0;	Mismatches 56;	
QY	194	ccttgagctcaactcagcagcgtggaataggtgttgatcacatcaatgaagtactctga	253
Db	52	ccgtgtccgccagcagctgaagtggtggagacagactcatcaacatcagcagtcaccctgt	111
QY	254	gcacgtccagaagaaggagcagttccctcgttgaagaagatccctcaagaccctcagctcg	313
Db	112	acggcagacagcagcaggaagccctatctcatcaaaaggcctgtgtccgcacatctgaagctca	171
QY	314	tagtgcagctga 327	
Db	172	cggtcagcaggtga 185	
RESULT 14			
LOCUS	AI825748/c	458 bp	MRNA
DEFINITION	wb969008.x1 NCL CGAP_Gc6 Homo sapiens CDNA IMAGE:2310903 3', mRNA sequence.		

```

ACCESSION      A1825748
VERSION        A1825748.1  GI:5446419
KEYWORDS
SOURCE
ORGANISM       human.
                Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS        1 (bases 1 to 458)
TITLE          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
                National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgaps-email.nih.gov
                Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
                R. Emmert-Buck, M.D., Ph.D.
                CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
                Bonaldo, Ph.D.
                CDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                www-bio.llnl.gov/bdrrp/image/image.html
                Insert Length: 1112 Std Error: 0.00
                Seq primer: -40UP from GIBCO
                High quality sequence stop: 455.
                Location/Qualifiers
                1..458
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2310903"
                /clone_11p="NCI-CGAP-GC6"
                /tissue_type="Pooled germ cell tumors"
                /lab_host="DH10B"
                /note="Vector: pRT3D-Pac (Pharmacia) with a modified
                polylinker. Site_1: Not I; Site_2: Eco RI; plasmid DNA
                from the normalized library NCI-CGAP clone G4 was prepared, and
                as circles were made in vitro. Following HAP purification,
                this DNA was used as tracer in a subtractive hybridization
                reaction. The driver was PCR-amplified cDNAs from a pool
                of 5,000 clones made from the same library (cloneds
                1257096-1258631, 1469064-1470963, and 1475582-1476743).
                Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT     147 a          95 c          93 g          119 t          4 others
ORIGIN
Query Match           12.9%; Score 42.2; DB 10; Length 458;
Best Local Similarity 82.5%; Pred. No. 0.16;
Matches 47; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 271 gcaattcccttgtaaagatccctcaagacccctcagctgtagtcgcagttga 327
      |||||||
DB 458 GCGATTCCCTGTGTAAGAGATCCTTNCACAGACCCTCAGCTGTGTCGCGATTGA 402
      |||||||

RESULT 15
AM356254          379 bp      mRNA          EST          25-APR-2001
DEFINITION       38503 MARC 2B0V Bos taurus CDNA 5', mRNA sequence.
ACCESSION        AM356254
VERSION          AM356254.1  GI:6860260
KEYWORDS
SOURCE
ORGANISM         Bos taurus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
                Bovidae; Bovinae; Bos.
                1 (bases 1 to 379)
                Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
                Caass,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett
                G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Shitko-McKown,C.G.,
                Pereira,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and

```

Search completed: January 31, 2002, 19:16:11
Job time: 1959 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 31, 2002, 19:45:37 ; Search time 1599.5 Seconds
(without alignments)
3372.665 Million cell updates/sec

Title: US-09-641-831-3
Perfect score: 327
Sequence: 1 atgatgggaccactgaaga.....ggctggtagtgcgagttga 327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_om:*

20: em_or:*

21: em_ov:*

22: em_pat:*

23: em_ph:*

24: em_pl:*

25: em_ro:*

26: em_sts:*

27: em_sy:*

28: em_un:*

29: em_vi:*

30: em_htgo_hum:*

31: em_htgo_inv:*

32: em_htgo_rod:*

33: em_htg_hum:*

34: em_htg_inv:*

35: em_htg_rod:*

36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	327	100.0	327	6	AX088025	Sequence
2	323	98.8	402	6	AX088027	Sequence
3	323	98.8	573	6	AX088023	Sequence
4	209	63.9	6423	10	AF199421	Mus muscu
5	170	52.0	163535	2	AC046189	Homo sapi
6	157.4	48.1	158375	2	AP002859	Homo sapi
7	97.4	29.8	7445	9	HSAPXL	H. sapiens A
8	82.8	25.3	6014	9	AB033028	Homo sapi
9	65	19.9	196542	9	AC002365	Homo sapi
10	45.8	14.0	156413	9	HSAL19E20	Human DNA
11	44.8	13.7	39213	9	AC003047	Homo sapi
12	44.8	13.7	44400	9	AC090481	Homo sapi
c	43.6	13.3	173510	9	AL359272	Human DNA
13	37.8	11.6	2111	9	S39392	Homo sapien
14	37.8	11.6	3984	9	HUNCAP	Human prote
15	36.8	11.3	1485	10	AF002283	Mus muscu
16	36.8	11.3	329709	1	AP002997	Mesorhizo
17	36.6	11.2	3220	9	BC008741	Homo sapi
18	36.6	11.2	3287	9	AF061258	Homo sapi
19	36.2	11.1	1696	10	AF095585	Rattus norv
20	35.8	10.9	1896	10	RNU48247	Homo sapi
21	35.8	10.9	134482	9	AC007335	Homo sapi
22	35.8	10.9	160460	2	AC020766	Homo sapi
c	35.6	10.9	112027	9	AC007006	Homo sapi
23	35.6	10.9	187727	2	AC064806	Homo sapi
24	34.8	10.6	1282	5	SSA308531	Salmo sal
25	34.8	10.6	147197	2	AC091912	Homo sapi
c	34.8	10.6	171280	2	AC024186	Homo sapi
26	34.8	10.6	183349	2	AC027501	Homo sapi
27	34.8	10.6	237308	2	AC009593	Homo sapi
c	34.6	10.6	1567	9	AF345903	Homo sapi
28	34.6	10.6	1669	9	AF345904	Homo sapi
29	34.6	10.6	1686	9	AF345906	Homo sapi
30	34.6	10.6	4542	3	AB036856	Ciona int
31	34.4	10.5	1607	9	HSAL13766	Homo sapi
32	34.2	10.5	1041	10	AB016587	Mus muscu
33	34.2	10.5	1191	10	RNRIT18	R. norvegicu
34	34.2	10.5	1196	10	AB016588	Mus muscu
35	34.2	10.5	1874	10	AB016586	Mus muscu
36	34.2	10.5	3619	9	BC010674	Homo sapi
37	34.2	10.5	3643	9	HUMPTYPH	Human prote
38	34.2	10.5	5121	10	AF074960	Mus muscu
39	34.2	10.5	181850	2	AL354983	Homo sapi
c	34.2	10.4	968	9	HUMBRP1	Human mRNA
40	34	10.4	4680	1	AF285084S2	Salmonell
41	34	10.4	4680	1	AF285085	Salmonell

ALIGNMENTS

RESULT 1	AX088025	Sequence 3	327 bp	DNA	PAT	17-MAR-2001
LOCUS	AX088025	Sequence 3	327 bp	DNA		
DEFINITION	AX088025	Sequence 3	327 bp	DNA		
ACCESSION	AX088025	Sequence 3	327 bp	DNA		
VERSION	AX088025.1	GI:13396952				
KEYWORDS	human					
SOURCE	human					
ORGANISM	Human sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
TITLE	1 (bases 1 to 327)					
JOURNAL	Turner, C.A., Zambrowicz, B., Friedrich, G., Nehls, M. and Sands, A.T.					
FEATURES	Sequence derived from a human mammary gland cdna library					
source	Patent: WO 0114422-A 3 01-MAR-2001;					
	Lexicon Genetics Incorporated (US)					
	Location/Qualifiers					
	1..327					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					

BASE COUNT 91 a 76 c 96 g 64 t
ORIGIN

Query Match 100.0%; Score 327; DB 6; Length 327;
Best Local Similarity 100.0%; Pred. No. 2,8e-88;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgatgagaccactgaagactcccaagcctagtccacattaaacttaacagcgc 60
DB 1 ATGATGAGGACCACTGAAGACTTCCACAGCCTAGTCCACATTAACTTAACAGGCGC 60
OY 61 accaaggagagatattatctgagagcattccctgagagagagcctccctgggtttt 120
DB 61 ACCAAGGAGAGATATTATCTGAGGCAATCTCTGAGGAGAGACTCCCTGGGGTTTT 120
OY 121 actctaaaggctgagcctgagcagcaggaagaaacattatccttaaggtcgaagaagg 180
DB 121 ACTCTAAAGGCTGAGCCTGAGCAGGAGAACCATTAATCATCTTAAGGTGGAAGAGGG 180
OY 181 ggcgaagcagacccctgagcctcaaacctgagcagcagcagcagcagcagcagcagc 240
DB 181 GGCgaagcagacccctgagcctcaaacctgagcagcagcagcagcagcagcagcagc 240
OY 241 gagtgactctgagcagcctcagaagaagagcagctccctggtgaaagatcctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGCAGTTCCCTGCTGAAGAGATCTTACAG 300
OY 301 accctcagcctgtagtgagcagttga 327
DB 301 ACCCTCAGGCTGTAGTGCGCAGTTGA 327

RESULT 2
AX088027 402 bp DNA PAT 17-MAR-2001
LOCUS AX088027 Sequence 5 from Patent WO0114422.
ACCESSION AX088027.1 GI:13396953
VERSION
KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 5 01-MAR-2001;
Lexicon Genetics Incorporated (US)

FEATURES
Location/Qualifiers
1..402
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 110 a 91 c 117 g 84 t
ORIGIN

Query Match 98.8%; Score 323; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 4.7e-87;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgatgagaccactgaagactccacaagcctagtccacattaaacttaacagcgc 60
DB 1 ATGATGAGGACCACTGAAGACTTCCACAGCCTAGTCCACATTAACTTAACAGGCGC 60
OY 61 accaaggagagatattatctgagagcattcctgagagagagcctccctgggtttt 120
DB 61 ACCAAGGAGAGATATTATCTGAGGCAATCTGAGGAGAGAGCCTCCCTGGGGTTTT 120
OY 121 actctaaaggctgagcctgagcagcaggaagaaacattatccttaaggtcgaagaagg 180
DB 121 ACTCTAAAGGCTGAGCCTGAGCAGGAGAACCATTAATCATCTTAAGGTGGAAGAGGG 180

OY 181 ggcgaagcagacaccctgagcctcccaactcagagcctgagagatgaggttgacatcaat 240
DB 181 GGCgaagcagacaccctgagcctcccaactcagagcctgagagatgaggttgacatcaat 240
OY 241 gagtgactctgagcagcctcagaagaagagcagctccctggtgaaagatcctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGCAGTTCCCTGCTGAAGAGATCTTACAG 300
OY 301 accctcagcctgtagtgagcag 323
DB 301 ACCCTCAGGCTGTAGTGCGCAG 323

RESULT 3
AX088023 573 bp DNA PAT 17-MAR-2001
LOCUS AX088023 Sequence 1 from Patent WO0114422.
ACCESSION AX088023
VERSION AX088023.1 GI:13396951
KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 573)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 1 01-MAR-2001;
Lexicon Genetics Incorporated (US)

FEATURES
Location/Qualifiers
1..573
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 151 a 152 c 152 g 118 t
ORIGIN

Query Match 98.8%; Score 323; DB 6; Length 573;
Best Local Similarity 100.0%; Pred. No. 4.8e-87;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgatgagaccactgaagactccacaagcctagtccacattaaacttaacagcgc 60
DB 1 ATGATGAGGACCACTGAAGACTTCCACAGCCTAGTCCACATTAACTTAACAGGCGC 60
OY 61 accaaggagagatattatctgagagcattcctgagagagagcctccctgggtttt 120
DB 61 ACCAAGGAGAGATATTATCTGAGGCAATCTGAGGAGAGAGCCTCCCTGGGGTTTT 120
OY 121 actctaaaggctgagcctgagcagcaggaagaaacattatccttaaggtcgaagaagg 180
DB 121 ACTCTAAAGGCTGAGCCTGAGCAGGAGAACCATTAATCATCTTAAGGTGGAAGAGGG 180
OY 181 ggcgaagcagacccctgagcctcaaacctgagcagcagcagcagcagcagcagcagc 240
DB 181 GGCgaagcagacccctgagcctcaaacctgagcagcagcagcagcagcagcagcagc 240
OY 241 gagtgactctgagcagcctcagaagaagagcagctccctggtgaaagatcctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGCAGTTCCCTGCTGAAGAGATCTTACAG 300
OY 301 accctcagcctgtagtgagcag 323
DB 301 ACCCTCAGGCTGTAGTGCGCAG 323

RESULT 4
AF199421 6423 bp mRNA ROD 13-DEC-1999
LOCUS AF199421 Mms musculus p2f domain actin binding protein Shroom mRNA, complete
DEFINITION
ACCESSION AF199421
VERSION AF199421.1 GI:6467989


```

----- Project Information
Center project name: 19512
Center clone name: 356.M17
----- Summary Statistics
Sequencing vector: M13; W77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 148971 bases at least Q40
Consensus quality: 156340 bases at least Q30
Consensus quality: 159016 bases at least Q20
Insert size: 181000; agarose-fp
Insert size: 160935; sum-of-ctrls
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2095: contig of 2095 bp in length
2096 2195: gap of 100 bp
2196 3520: contig of 1325 bp in length
3521 3620: gap of 100 bp
3621 4777: contig of 1157 bp in length
4778 4877: gap of 100 bp
4878 6660: contig of 1783 bp in length
6661 6760: gap of 100 bp
6761 8530: contig of 1770 bp in length
8531 8630: gap of 100 bp
8631 11149: contig of 2519 bp in length
11150 11249: gap of 100 bp
11250 15298: contig of 4049 bp in length
15299 15398: gap of 100 bp
15399 18533: contig of 3135 bp in length
18534 18633: gap of 100 bp
18634 22044: contig of 3411 bp in length
22045 22144: gap of 100 bp
22145 25578: contig of 3434 bp in length
25579 25678: gap of 100 bp
25679 29571: contig of 3893 bp in length
29572 29671: gap of 100 bp
29672 34405: contig of 4734 bp in length
34406 34505: gap of 100 bp
34506 37415: contig of 2910 bp in length
37416 37515: gap of 100 bp
37516 44162: contig of 6647 bp in length
44163 44262: gap of 100 bp
44263 48837: contig of 4575 bp in length
48838 48937: gap of 100 bp
48938 55115: contig of 6178 bp in length
55116 55215: gap of 100 bp
55216 62194: contig of 6979 bp in length
62195 62294: gap of 100 bp
62295 69547: contig of 7253 bp in length
69548 69647: gap of 100 bp
69648 78050: contig of 8403 bp in length
78051 78150: gap of 100 bp
78151 88041: contig of 9891 bp in length
88042 88141: gap of 100 bp
88142 98560: contig of 10419 bp in length
98561 98660: gap of 100 bp
98661 108146: contig of 9866 bp in length
108147 108246: gap of 100 bp
108247 118855: contig of 10609 bp in length
118856 118955: gap of 100 bp
118956 128543: contig of 9588 bp in length
128544 128643: gap of 100 bp
128644 142647: contig of 14004 bp in length
142648 142747: gap of 100 bp

```

```

FEATURES
source
* 142748 153569: contig of 10822 bp in length
* 153570 153669: gap of 100 bp
* 153670 163535: contig of 9866 bp in length.
Location/Qualifiers
1. 163535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-356M17"
/clone_11D="RPC1-11 Human Male BAC"
1. 2095
/note="assembly-fragment
clone_end:5P6
vector_side:left"
2196. 3520
/note="assembly-fragment"
3621. 4777
/note="assembly-fragment"
4878. 6660
/note="assembly-fragment"
6761. 8530
/note="assembly-fragment"
8631. 11149
/note="assembly-fragment"
11250. 15298
/note="assembly-fragment"
15399. 18533
/note="assembly-fragment"
18634. 22044
/note="assembly-fragment"
22145. 25578
/note="assembly-fragment"
25679. 29571
/note="assembly-fragment"
29672. 34405
/note="assembly-fragment"
34506. 37415
/note="assembly-fragment"
37516. 44162
/note="assembly-fragment"
44263. 48837
/note="assembly-fragment"
48938. 55115
/note="assembly-fragment"
55216. 62194
/note="assembly-fragment"
62295. 69547
/note="assembly-fragment"
69648. 78050
/note="assembly-fragment"
78151. 88041
/note="assembly-fragment"
88142. 98560
/note="assembly-fragment"
98661. 108146
/note="assembly-fragment"
108247. 118855
/note="assembly-fragment"
118956. 128543
/note="assembly-fragment"
128644. 142647
/note="assembly-fragment"
142748. 153569
/note="assembly-fragment"
153670. 163535
/note="assembly-fragment
clone_end:T7
vector_side:right"
BASE COUNT 48438 a 32205 c 31514 g 48770 t 2608 others
ORIGIN

```

Query Match	52.0%	Score 170;	DB 2;	Length 163535;
Best Local Similarity	100.0%;	Pred. No. 1.5e-40;		
Matches 170;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	atgatgagaccactgaagacttccacaagcctagtcgacacattaaactcaacacggcc 60			
Db 62580	ATGATGAGACCACTGAAGACTTCCACAGCCTAGTGCACATTAACACTAACACGGCC 62639			
QY 61	accgaaggaaagtacattatctctggaggcattctctggaggaggagggcctctctggggtttt 120			
Db 62640	ACCAAGGAAGTACATTATCTCTGGAGGATTCCTCTGGAGGAGGAGCTCCCTCGGGGTTTT 62699			
QY 121	actctaaagggtggcctggagcagcgagacacattatcatctctaaagt 170			
Db 62700	ACTCTAAAGGGTGGCCTGGAGCAGCGAGAACCATTAATCATCTCTAAGGT 62749			
RESULT 6				
AP002859	AP002859 156375 bp DNA HTG 29-AUG-2000			
LOCUS	Homo sapiens chromosome 4 clone 225E15 map 4q16-q20, *** SEQUENCING			
DEFINITION	IN PROGRESS ***, 15 unordered pieces.			
ACCESSION	AP002859			
VERSION	AP002859.1 GI:9955382			
KEYWORDS	HTG; HTGS; PHASE1			
SOURCE	Homo sapiens DNA, clone:225E15.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1. (bases 1 to 156375)			
JOURNAL	Tsai, S.F.			
COMMENT	Direct Submission Submitted (28-AUG-2000) Shih-Feng Tsai, National Yang-Ming University, Institute of Genetics; 155 Li-Rong St. Section 2, Peitou, Taipei, Taiwan 11221, Republic of China (E-mail: ymptsai@ym.edu.tw, URL: http://genome.ym.edu.tw/, Tel: 886-2-28267043, Fax: 886-2-28264930) These sequences are draft human sequences, not finished sequences. These sequences are unordered pieces. And gaps between the contigs of the same clone are represented as 100 N. * NOTE: This is a 'working draft' sequence. It currently * consists of 15 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 ... 21185: contig of 21185 bp in length * 21186 21285: gap of 100 bp * 21286 22928: contig of 1643 bp in length * 22929 23028: gap of 100 bp * 23029 26285: contig of 3257 bp in length * 26286 26385: gap of 100 bp * 26386 30606: contig of 4221 bp in length * 30607 30706: gap of 100 bp * 30707 36924: contig of 6218 bp in length * 36925 37024: gap of 100 bp * 37025 40953: contig of 3929 bp in length * 40954 41053: gap of 100 bp * 41054 47843: contig of 6790 bp in length * 47844 47943: gap of 100 bp * 47944 57174: contig of 9331 bp in length * 57175 57274: gap of 100 bp * 57275 64994: contig of 7720 bp in length * 64995 65094: gap of 100 bp * 65095 76523: contig of 11429 bp in length * 76524 76623: gap of 100 bp * 76624 88306: contig of 11683 bp in length * 88307 88406: gap of 100 bp * 88407 101162: contig of 12756 bp in length * 101163 101262: gap of 100 bp * 101263 115436: contig of 14174 bp in length			

Fri Feb. 1 08:56:28 2002

```

RESULT 9
AC002365 196542 bp DNA PRI 22-AUG-1998
LOCUS Homo sapiens chromosome X clone UI77G4, U152H5, U168D5, U17A6,
DEFINITION U172D6, and U186B3 from Xp22, complete sequence.
ACCESSION AC002365 U82630 U83511
VERSION AC002365.1 GI:2358015
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 196542)
AUTHORS Muzny,D., Aronson,A.D., Adams,C., Brundage,E., Bunac,C.,
Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S.,
Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M.,
Gorrell,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S.,
Kampal,R., Karpach,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O.,
Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,
Rashid,M.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,
Simon,M., Stovall,K., Timms,K.M., Todd,J., Vo,O., Williamson,A.,
Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 196542)
AUTHORS Chiu,M.W.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 196542)
AUTHORS Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Gonzalez,O.,
Haywood,M., Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L.,
Shen,H., Worley,K., Chen,E., Forcum,J., Aronson,A.D., Chiu,M.W.,
Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and
Gibbs,R.A.
Direct Submission
Submitted (20-AUG-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 196542)
AUTHORS Chiu,M.W.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA
5 (bases 1 to 196542)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 4, 1997 this sequence version replaced gi:1766071 gi:1773380
gi:2289024.
COMMENT Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.
The repeat regions shown were identified using RepeatMasker by
Adrian Smit.
Sequence similarities were identified using Powerblast by Jinghui
Zhang.
Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
LOCATION/Qualifiers
1. 196542
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source
/clone="U177G4"
/chromosome="X"
/clone_lib="Livermore Chromosome X Cosmid Library
LLOXNC01"
/map="Xp22"
/complement(760..1057)
/rpt_family="AluJb"
1853..2194
/rpt_family="AluSq"
2305..2584
/rpt_family="AluSx"
3418..3630
/rpt_family="MER20"
complement(3806..4096)
/rpt_family="AluSx"
4138..4537
/rpt_family="L1MB2"
complement(4549..4841)
/rpt_family="AluJo"
complement(5421..5528)
/rpt_family="MER58B"
complement(7001..7132)
/rpt_family="FAM"
7773..7903
/rpt_family="AluJ"
7910..8215
/rpt_family="AluYa5"
8224..8401
/rpt_family="AluJ"
complement(8932..9232)
/rpt_family="AluY"
complement(9458..9481)
/rpt_family="POLY_A"
11579..11618
/rpt_family="AT-rich"
complement(11619..11916)
/rpt_family="AluSg"
complement(11917..12252)
/rpt_family="AluJb"
complement(12328..12633)
/rpt_family="AluSx"
complement(12665..13117)
/rpt_family="L1MB3"
complement(13236..13267)
/rpt_family="L1MA1"
13323..13604
/rpt_family="AluJb"
13983..14378
/note="Similar to sequence with GenBank Accession number
236840"
repeat_region
15384..15665
/rpt_family="MER58B"
15656..15956
/rpt_family="AluSg"
16091..16660
/rpt_family="L1MB7"
join(<17023..17176,34350..34481,37731..40071,41563..41663,
75548..76243,80507..81058,82568..82739,88014..88286,
90044..>92814)
90044..>92814)
/gene="HSAPXL"
/product="APXL"
join(<17023..17176,34350..34481,37731..40071,41563..41663,
75548..76243,80507..81058,82568..82739,88014..88286,
90044..>92814)
/gene="HSAPXL"
<17023..>92814
/gene="HSAPXL"
join(<17023..17176,34350..34481,37731..40071,41563..41663,
75548..76243,80507..81058,82568..82739,88014..88286,
90044..>90310)
/gene="HSAPXL"
/note="human homolog of Xenopus laevis APX gene, similar
to sequence encoded by GenBank Accession Number X83543"

```

```

/codon_start=3
/product="APX1"
/protein_id="AAC02592.1"
/db_xref="GI:1173381"

/translation="TEGSGAAAVDKLLAGDEIVGINDIGSGFROEALCLVKGSHKT
LKLVAKRSELGMPHSMWHTATYSDSHPELAASTPTSTGSGCWGRRHAASSSHDL
SWMOTNLQRLDHSLSGYSKIDSHPSLSTVAKNSIDHLAGSGVSSGSGFSG
SSPHDLSLADTSSAKNIEIVGLWELAVPQKQAQAAGPOSEKSLCTFPYVAV
GDSKGRPEPVEAKMLAAGRNENGVVVPQKKAAGSPPEPPPLRDSFATKRA
HEKAGCVFEAAAHQHTLAAQPGDREPLDRMRSNHLPSLCKSGSGECPC
EAHDSMPKSGSRLQASLSSDVRPSPSHGRPHLSPHSLCADISGQEP
GASFGNDSPVQVSGSCDOKRSGSCMGCPREYQGDLOALNANGMSPHAGLE
SLPPTTQVSPRHLPDQBPDPAREGRETLDKACRAGAPPPHASEAAS
RLAASITWAGSESRICPOTPLPHSTIDGKRRPSSPESDTSAGDPPHASEAAS
KSDHFAATLLENYLQMAKATQKRSRTVLTAAAEADGMRNALGGGTQGLAGT
RYDDELAQAQVAVLAFSFKRDLDPNVDTPESLLEHMGDNGVLEPMWGLAG
RKHLEPPTPIGGRRTALQOKLSIPKEMKNEVLTAGTSPHONPTSEDTVGL
DRMKEPFTKPVQPAQKQALHGIPEKPEPRTAGTCGTPGWSRTTSLGSL
AHSAEKAAGSPLDRLRTGAEVQSWKEDRKLKRSRGRHSAADLTVLSIDPQEE
POVHNGSGSSPDHRTKQDAVSLKROAGDPEPEPLPSAVRABEOSNPRQADAL
CRBESPSOCHPSPQAKANPPTSELSHCAGAPDLBREGCGAATLPPDXYRSESTPT
ADLQKRSQFSPSLHAGQDSWVYSSLSLKRAQPPRPKREKRETRATDGAAD
PVGVLQPPPTSPADLVAVLSHSPSVSSAQPDPTKAVRCRSGAHVSDGA
PRPEALLERQOHNLQATVETSRSPQAPKLDKRPDLIDESTRTERVNQA
DNNTVYAPVCIKVHSSQEKRSRSLQCAPAPRLPGLKGLDITLSDSTFQ
FCLVTCGAGAEENHAKROPEPLQGVPEKDRCTSPGYSYAKAAEKTYEKL
ELAEIVGGRKSLADLPDPEVKTITMDLMEGIFQDEHLEAGQORRKLRLIP
STKEERKEPVPAAVSLATSTSYSSSAPKALLKMRLOQCHHEDPSDDHDL
SYKQKVELLETSKRLQVLRARRSLSLSDQVANTYUAEVLAUYKQVCSPEDFRNF
IGDLDKVELLETSGLGVAEVALNNLNDQVSGPGRSLERKOVLLQDHAELEKL
INDRRERIVFDILANTYLSSESLADYEHPYKMSALLIQEPLERDKIHLGEOLKCL
SILQPEGRK"
```

```
repeat_region complement(15976, .18074)
/rfp_family="HY3"
18162, .18473
repeat_region /rfp_family="AlubB"
19536, .19717
/rfp_family="MER21A"
complement(19798, .19937)
/rfp_family="11MB3"
complement(20181, .20327)
/rfp_family="GRAM"
complement(22367, .22401)
/rfp_family="AT_rich"
22406, .22461
/rfp_family="(GCAN)n"
complement(22462, .22627)
/rfp_family="AluJo"
complement(22638, .22933)
/rfp_family="AluSx"
complement(26881, .26957)
/rfp_family="MER5A"
complement(27726, .27762)
/rfp_family="POLY_A"
complement(27876, .27977)
/rfp_family="AluJo/GRAM"
complement(28048, .28069)
/rfp_family="AT_rich"
```

Query Match	19.98;	Score 65;	DB 9;	Length 196542;
Best Local Similarity	64.1%;	Pred. No. 8.7e-09;		
Matches 98;	Conservative 0;	Mismatches 55;	Indels 0;	Gaps 0

OY 175 gaaggggcaaaacaacacccttgctgcgaacatcgtagcgttggagtgtgaattcac 234
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17028 GAGAGGGCGCTAAACCCCGGCGGTCCACAAGTACTGCTGTGAGATGGAATCCTGC 17087

OY 235 atcaatgagtgtaaccttcgacgactccagaagaaggagcagttccctcgtgtgaagatcc 294
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17088 ATCAATGACACTTGTCCTCCAGAGGTTTAGACAGGAAGCATTTCTCGTGAAGGGATCC 17147

OY 295 tacaagacccctcagctggttaatgtgcagtttga 327
| | | | | | | | | | | | | | | | | | | | | | | |

Db 17148 CATAAGACCCCTGAAGCTGCTGTCACAAAGGTA 17180

Accession	Definition	LOCUS	Result 10
U113165	Human DNA sequence from clone RP11-119E20 on chromosome Xp11.21-11.23 Contains part of the gene for KIAA1202 protein, ESTs STS and GSS, complete sequence.	HSAL19E20 156413 bp	14-SEP-2000

ACCESSION	AL121865
VERSION	AL121865.7
KEYWORDS	HTG; KIAA1202.
SOURCE	human.

ORGANISM	REFERENCE AUTHORS TITLE JOURNAL
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
1 (bases 1 to 156413)	
Wilson, S.	
Direct Submission	
Submitted (27-JUL-2000)	Sanger Centre, Hinxton, Cambridgeshire,

COMMENT
On Jul 8, 2000 this sequence version replaced gi:18977865.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure'

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information from the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chrx> RP11-119E20 is from the library RPC1-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PBAC3.6> This sequence is the entire insert of clone RP11-119E20.

```

FEATURES
    source
        Location/Qualifiers
            1..156413
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="X"
                /map="p11-21-11.23"
                /clone="RP11-119E20"
                /clone_1lb="RPC1-11.1"
            8..593
                /note="match: GSS: Em:AQ350374"
            1965..2081
                /note="L2 repeat: matches 2602..2708 of consensus"
            2374..2464
                /note="L2 repeat: matches 2610..2686 of consensus"
            2607..2755
                /note="MIR repeat: matches 27..176 of consensus"
            join<3161..3312,60790..60924,63430..65920,71423..71484,
            90914..91717,96285..96465,100563..100832,102134..103400
                /gene="bAl19E20.1"
            /note="match: cDNAs: Em:AB033028 Em:X63543 Em:AB040914"
            /evidence=not_experimental
            /product="bAl19E20.1 (KIAA1202 protein)"
            3161..103400
                /gene="bAl19E20.1"

```



```

DB      3276 AAGGCTCCTCCGATTCACGATGATGACGAGGTAA 3316
RESULT  11
AC003047 39213 bp DNA PRI 30-MAY-1998
LOCUS     Homo sapiens Xp22 Cosmid U152D7 (Lawrence Livermore human cosmid
DEFINITION library) complete sequence.
ACCESSION AC003047
VERSION    AC003047.1 GI:3169149
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 39213)
AUTHORS   Muzny D., Areson A.D., Adams C., Brundage E., Bunac C.,
            Carvelli K., Chacko J., Chen J., Di W., Ding Y., Dugan S.,
            Durbin J., Forcum J., Ganes R., Garcia C., Goodman M.,
            Gorrell J.H., Haywood M., Hernandez J., Jackson L., Jin S.,
            Kampal R., Karpathy S., Kovar C., Leal B., Li Y., Lichtarge O.,
            Liu W., Logan O., Lu J., Ly T., Martinez C., Oswal G., Perez L.,
            Rashid N.D., Rowland K., Savage L., Scherer S.E., Shen H.,
            Simon M., Stovall K., Timms K.M., Todd J., Vo O., Williamson A.,
            Worley K.C., Yu W., Chinault C., Nelson D. and Gibbs R.A.
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE 2 (bases 1 to 39213)
AUTHORS   Worley K.C.
TITLE      Direct Submission
JOURNAL    Submitted (03-NOV-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 39213)
AUTHORS   Worley K.C.
TITLE      Direct Submission
JOURNAL    Submitted (30-MAY-1998) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT    On May 30, 1998 this sequence version replaced gi:3108011.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.

FEATURES             source
    source
        1..39213
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="U152D7"
            /chromosome="X"
            /clone_lib="Lawrence Livermore human cosmid library"
            /map="Xp22"
            complement(9..118)
            /rpt_family="MLT1A1"
            119..438
            /rpt_family="L1MB6"
            complement(447..760)
            /rpt_family="AluYo"
            repeat_region
            complement(968..1179)
            /rpt_family="MIR"
            1308..1602
            /rpt_family="AluYb"
            repeat_region

```

```

repeat_region complement(1710..1835)
            /rpt_family="FLAM_A"
repeat_region complement(2789..3087)
            /rpt_family="AluSx"
repeat_region complement(3095..3126)
            /rpt_family="AT_rich"
repeat_region complement(3245..3375)
            /rpt_family="MERSA"
            /rpt_family="MERSA"
            3622..3712
            /rpt_family="(GGA)n"
repeat_region complement(3762..3788)
            /rpt_family="AT_rich"
repeat_region complement(3789..4087)
            /rpt_family="AluSx"
            4514..4801
            /rpt_family="AluSx"
repeat_region 4804..5096
            /rpt_family="AluSx"
            complement(5643..5927)
            /rpt_family="AluYo"
            5928..6181
            /rpt_family="AluYo"
            6505..6701
            /rpt_family="MERSB"
            6715..6995
            /rpt_family="AluYb"
repeat_region 7006..7092
            /rpt_family="MERSB"
            7191..7227
            /rpt_family="(CA)n"
            7428..7724
            /rpt_family="AluSx"
            8196..8266
            /note="low coverage"
            complement(8304..8355)
            /rpt_family="GC_rich"
            8705..8789
            /rpt_family="(GGA)n"
            9136..9429
            /rpt_family="AluSx"
            9430..9580
            /rpt_family="AluYb"
            9588..9625
            /rpt_family="(CA)n"
            9627..9663
            /rpt_family="AT_rich"
            10920..11037
            /rpt_family="AluYb"
            11146..11196
            /note="low coverage"
            11203..11335
            /rpt_family="FRAM"
            complement(11351..11652)
            /rpt_family="AluSx"
            complement(12228..12254)
            /rpt_family="AT_rich"
            complement(12257..12556)
            /rpt_family="AluSx"
            12557..12590
            /rpt_family="AT_rich"
            12704..12994
            /rpt_family="AluYo"
            12995..13020
            /rpt_family="(GA)n"
            13023..13093
            /rpt_family="MERSA"
            13667..13854
            /rpt_family="AluSx"
            complement(14127..14233)
            /rpt_family="MERSA"
            complement(14368..14670)
            /rpt_family="AluSx"
            15204..15494

```

[illegible]

flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Contig Length: 44400
Phrap values in estimate: 44245
Average error rate (BCM-Phrap estimate): 7.63605e-07
Fraction of Phrap values less than 40 : 0.00101706
Number of consensus changing edits: 10
Number of N's in consensus : 0
```

```
----- Consensus changing edits -----
Position Original+Context Edited+Context
12913 tcaattctat(n)tatattgatt tcaattctat(t)tatattgatt
20039 ccattccaat(n)tgaaatgatt ccattccaat(t)tgaaatgatt
38071 ggaaggaattg(n)tgctcttacc ggaaggaattg(t)tgctcttacc
38096 tggatgcag(n)atttgagctg tggatgcag(t)atttgagctg
44395 catctctg(n)nnnnngcgcc catctctg(g)nnnnngcgcc
44396 atctctg(n)nnnnngcgcc atctctg(g)nnnnngcgcc
44397 tctctg(n)nnnnngcgcc tctctg(g)nnnnngcgcc
44398 ctctg(n)nnnnngcgcc ctctg(g)nnnnngcgcc
44399 tctg(n)nnnnngcgcc tctg(g)nnnnngcgcc
44400 ctg(n)nnnnngcgcc ctg(g)nnnnngcgcc
```

----- Distribution of Quality < 40 Bases -----

```
501
451
401
351
301
251
201
151
101
51
01
5 10 15 20 25 30 35 40
Phrap Value Range
```

FEATURES

```
Source
Version: 1.01 gxf.
Location/Qualifiers
1. 44400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Xp"
/clone="RP11-8H7"
complement(1..2004)
/note="Overlaps bases 2004..1 of clone AC003036"
/function="Overlaps with adjacent clone AC003036"
```

```
repeat_region 948..1001
repeat_region /rpt_family="MLTJ"
repeat_region 1002..1036
repeat_region /rpt_family="AluSq"
repeat_region 1374..1744
repeat_region /rpt_family="MLTJ"
repeat_region 2033..2946
repeat_region /rpt_family="LMB7"
repeat_region complement(2947..3287)
repeat_region /rpt_family="MLT1A1"
repeat_region 3288..3607
repeat_region /rpt_family="LMB7"
repeat_region complement(3627..3929)
repeat_region /rpt_family="AluJo"
repeat_region complement(4137..4348)
repeat_region /rpt_family="MIR"
repeat_region 4477..4779
repeat_region /rpt_family="AluDb"
repeat_region complement(4878..5004)
repeat_region /rpt_family="PLAM.A"
gene complement(join(5198..5292,6583..6692,11434..11743))
gene "H.sapiens mRNA (ocular albinism type 1 related)"
repeat_region /rpt_family="TTTA)n"
repeat_region 5879..5957
repeat_region /rpt_family="TTTA)n"
repeat_region complement(5958..6257)
repeat_region /rpt_family="AluX"
repeat_region 6258..6294
repeat_region /rpt_family="TTTA)n"
repeat_region complement(6399..6544)
repeat_region /rpt_family="MERSA"
STS 6502..7047
/standard_name="516/518"
/db_xref="GDB:591575"
repeat_region 6791..6881
repeat_region /rpt_family="GA-rich"
repeat_region complement(6950..7256)
repeat_region /rpt_family="AluX"
repeat_region 7685..7972
repeat_region /rpt_family="AluX"
repeat_region 7974..8267
repeat_region /rpt_family="AluX"
repeat_region 8345..8509
repeat_region /rpt_family="(CAATA)n"
repeat_region complement(8818..9098)
repeat_region /rpt_family="AluJo"
repeat_region 9099..9352
repeat_region /rpt_family="AluJo"
repeat_region 9353..9492
repeat_region /rpt_family="L1MC5"
repeat_region complement(9676..9885)
repeat_region /rpt_family="MERS8B"
repeat_region 9886..10166
repeat_region /rpt_family="AluDb"
repeat_region complement(10167..10258)
repeat_region /rpt_family="MERS8B"
STS 10294..10455
/standard_name="CA1-CA"
```

Query Match 13.7% Score 44.8 DB 9 Length 44400;
Best Local Similarity 69.3% Pred. No. 0.01;
Matches 61; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
QY 83 tggaggaattctctggaaggaaggaactccctgggggtttactctaaaggtggcctggagc 142
|||||
Db 32491 TGGAGGTGCACACTGAGCGCGCCCGGTGGGGCTTACCTGAAGGCGCGCGAGC 32550
|||||

QY 143 acggagaaacattaaatcatctctaagt 170
|||||
Db 32551 ACGCGAGCCGCTGCTCATCCACCAAGT 32578
|||||
```

RESULT 13
AL359272/c

LOCUS AL359272 173510 bp DNA 12-OCT-2000
DEFINITION Human DNA sequence from clone RP11-554P16 on chromosome X. Contains the first coding exon of the gene KIAA1202, STSs, GSSs and a CpG island, complete sequence.
ACCESSION AL359272
VERSION AL359272.9 GI:9650579
KEYWORDS HTG; CpG island; KIAA1202.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 173510)
TITLE Bird.C.
JOURNAL Direct Submission
COMMENT Submitted (28-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Aug 1, 2000 this sequence version replaced gi:9588592.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>
 RP11-554P16 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pBACe3.6
 This sequence is the entire insert of clone RP11-554P16.

FEATURES
 source
 Location/Qualifiers
 1..173510
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone="RP11-554P16"
 /clone_lib="RPCI-11.2"
 2..192
 /note="L1MEC repeat: matches 1913. .2098 of consensus"
 7..674
 /note="match: GSS: Em:AQ373815"
 195..511
 /note="AluSp repeat: matches 1. .312 of consensus"
 543..845
 /note="AluSx repeat: matches 1. .304 of consensus"
 950..1127
 /note="L1MEC repeat: matches 1968. .2136 of consensus"
 1190..2306
 /note="L1M4b repeat: matches -180. .888 of consensus"
 2425..3343
 /note="L1M4 repeat: matches 3078. .4092 of consensus"
 3344..3717
 /note="MSTD repeat: matches 1. .394 of consensus"
 3718..4412
 /note="L1M4 repeat: matches 4092. .4274 of consensus"
 4413..4452
 /note="THE1B repeat: matches 1. .42 of consensus"
 4453..4763
 /note="AluSx repeat: matches 1. .310 of consensus"

4764..5111
 /note="THE1B repeat: matches 42. .364 of consensus"
 5112..5774
 /note="L1M4 repeat: matches 4274. .5344 of consensus"
 5775..6080
 /note="AluSg repeat: matches 1. .307 of consensus"
 6081..6296
 /note="L1M4 repeat: matches 5344. .5581 of consensus"
 6299..6923
 /note="L1PA10 repeat: matches 5525. .6165 of consensus"
 7394..7652
 /note="L1M4 repeat: matches 2120. .2383 of consensus"
 7652..10378
 /note="L1PA4 repeat: matches 3414. .6142 of consensus"
 10369..10878
 /note="L1M4 repeat: matches 2350. .2879 of consensus"
 10875..11250
 /note="L1 repeat: matches 4741. .5131 of consensus"
 11247..11619
 /note="L1MA7 repeat: matches 5903. .6284 of consensus"
 11642..11764
 /note="L1M4 repeat: matches 5351. .5475 of consensus"
 11764..11943
 /note="L1 repeat: matches 3846. .4036 of consensus"
 12269..13359
 /note="L1MC3 repeat: matches 6659. .7739 of consensus"
 complement(13126..13490)
 /note="match: STS: Em:HS308ZG9"
 13430..13497
 /note="L1M4 repeat: matches 1995. .2062 of consensus"
 13498..13803
 /note="AluJo repeat: matches 3. .307 of consensus"
 13804..15085
 /note="L1M4 repeat: matches 720. .1995 of consensus"
 15191..15376
 /note="L1M4 repeat: matches 408. .602 of consensus"
 15585..16035
 /note="L1M4 repeat: matches -252. .225 of consensus"
 16058..16114
 /note="L12 repeat: matches 2686. .2745 of consensus"
 16128..16369
 /note="L1MA10 repeat: matches 5928. .6318 of consensus"
 17285..17533
 /note="match: GSS: Em:AQ134567"
 17455..17899
 /note="match: GSS: Em:AQ142071"
 17875..18018
 /note="MIR repeat: matches 96. .252 of consensus"
 18979..19286
 /note="AluDb repeat: matches 2. .308 of consensus"
 19391..19823
 /note="L12 repeat: matches 2284. .2708 of consensus"
 20320..20610
 /note="AluSx repeat: matches 1. .289 of consensus"
 21916..22121
 /note="match: STS: Em:G03829"
 23553..23862
 /note="MER33 repeat: matches 1. .307 of consensus"
 23946..24118
 /note="MIR repeat: matches 77. .259 of consensus"
 24356..24475
 /note="L12 repeat: matches 2624. .2750 of consensus"
 24929..25125
 /note="MIR repeat: matches 23. .256 of consensus"
 25251..25386
 /note="MER63 repeat: matches 2. .768 of consensus"
 25604..25656
 /note="L12 repeat: matches 2657. .2709 of consensus"
 26047..27589
 /note="L1MB8 repeat: matches 4591. .6159 of consensus"
 27590..27886
 /note="AluSx repeat: matches 1. .297 of consensus"
 27887..28214

```

repeat_region /note="LIMB8 repeat: matches 4271. .4591 of consensus"
28258. .29096 /note="LIMB8 repeat: matches 3264. .4157 of consensus"
repeat_region 29118. .29521 /note="MUTB repeat: matches 1. .383 of consensus"
repeat_region 29523. .30381 /note="LIMB8 repeat: matches 2496. .3231 of consensus"
repeat_region 30382. .30685 /note="LIMB8 repeat: matches 2. .299 of consensus"
repeat_region 30686. .31069 /note="LIMB8 repeat: matches 2110. .2496 of consensus"
repeat_region 31088. .31421 /note="LIMB8 repeat: matches 5820. .6156 of consensus"
repeat_region 31422. .31617 /note="LIMB8 repeat: matches 5374. .5552 of consensus"
repeat_region 31618. .31909 /note="LIMB8 repeat: matches 1. .294 of consensus"
repeat_region 31910. .31976 /note="LIMB8 repeat: matches 5304. .5374 of consensus"
repeat_region 31978. .32359 /note="LIMB8 repeat: matches 1951. .2330 of consensus"
repeat_region 32354. .33592 /note="LIMB8 repeat: matches 969. .2197 of consensus"
repeat_region 33935. .34790 /note="LIMB8 repeat: matches -266. .602 of consensus"
repeat_region 34856. .35219 /note="LIMB8 repeat: matches 1. .364 of consensus"
repeat_region 36043. .36159 /note="LIMB8 repeat: matches 2576. .2710 of consensus"
repeat_region 38455. .39057 /note="LIMB8 repeat: matches 2576. .2710 of consensus"
misc_feature /note="match: GSS: Em: A0008362"
41978. .42373 /note="match: STS: Em: L18363"
42056. .42087 /note="match: STS: Em: L18363"
repeat_region 42311. .42540 /note="LIMB8 repeat: matches 19. .262 of consensus"
repeat_region 42824. .42914 /note="LIMB8 repeat: matches 2606. .2709 of consensus"
repeat_region 43119. .43173 /note="LIMB8 repeat: matches 2687. .2742 of consensus"
repeat_region 43122. .43185 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region 44700. .44743 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region 45956. .46137 /note="LIMB8 repeat: matches 20. .199 of consensus"
repeat_region 46401. .46617 /note="LIMB8 repeat: matches 23. .262 of consensus"
repeat_region 46729. .47035 /note="LIMB8 repeat: matches 1. .309 of consensus"
repeat_region 47439. .47682 /note="LIMB8 repeat: matches 5666. .5914 of consensus"
repeat_region 48430. .48527 /note="LIMB8 repeat: matches 5666. .5914 of consensus"

Query Match 13.3%: Score 43.6; DB 9; Length 173510;
Best Local Similarity 67.8%; Pred. No. 0.026;
Matches 61; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```

RESULT 14
 LOCUS S39392 2111 bp mRNA PRI 05-MAR-2001
 DEFINITION Homo sapiens protein tyrosine phosphatase mRNA, partial cds.
 ACCESSION S39392

```

VERSION S39392.1 GI:250890
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2111)
AUTHORS Arimura, Y., Hinoda, Y., Itoh, F., Takekawa, M., Tsujisaki, M.,
Adachi, M., Inai, K. and Yachi, A.
TITLE CDNA cloning of new protein tyrosine phosphatases in the human
JOURNAL Tumor Biol. 13 (3), 180-186 (1992)
MEDLINE 92327504
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI gidsq 108079] from the original journal article.
This sequence comes from Fig. 3.
FEATURES
Source location/Qualifiers
1..2111
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="C16"
/tissue-type="colon"
<1..>2111
/note="Prase: This sequence comes from Fig. 3"
/codon_start=3
/product="protein tyrosine phosphatase"
/protein_id="AAB22439.2"
/db_xref="GI:13195754"
/translaction="OHSGIKOSEAESCTINARTLDFYGEVLAHSGRLHNLMLIGIA
SAGVAVRYKTCSEFYPMVNTLTKSPKRFTHOROKAORSHIYAFNNLRSCK
NMKSCVHHTFPQAKTLPOEKNTLSQYMKFGRNKNVNOYCKKVGWYMNFA
MRSLSEHETKSLPSPSPPTTPMWRSPRIHRIKPRHRSADNLNENYITETED
VFIYKQSLADPDSSEVSONRSPHOSLSNNPQSTLTOKSSVSPSNAGSCS
PDGVDOOLDFPHRYTKGSTEDASQYCDKNDGDSYLVLRTPEDGKFGNLS
GVDOKMPVVRINPESPADICIRLNGDQIVLNGDISEHTHDVYVAFIKARS
HSRELATVIRRAVRSFADPKSEDELNPPEALFPMCPGEGDTLESGMOLKARS
GTVLIOPOLYKRRKGLAIFPAKLPONDKRYRDVLPDYTVLNGEDNYINASY
MTEIPANTLVKRYLATOGSLPHTCAGQFQVYWDOKLSLIVLTLTTERGRKCHQYMP
DPPVYMNHGFHICQSEDCITAIYVSRMLVTNTOGSEHTVTHQYVAMPDHPID
SSDFEAFNVYRSLVDSSEVYVHCSAIGTGVLYMETAMCLTERNLPYLDIVR
KMDORAMVQTSQYKFCVCE"
BASE COUNT 594 a 528 c 510 g 479 t
ORIGIN

```

```

Query Match 11.6%: Score 37.8; DB 9; Length 2111;
Best Local Similarity 53.8%; Pred. No. 1.1;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 97 gaaggaagagccctccctgggtttactctaaaggggtgctgagacaggaacacat 156
DB 969 GATGAAGATGGAATAATTGGATTTAATCTTAAGGAGGATGGATCAAAAGATGCTCT 1028

QY 157 atcaatcctaaggtcgaagaagggaaggaagacacccctgaagcactgaagct 216
DB 1029 GTGGATTCGAAGATTAACCCAGAGTCACCTCGGACACCTGCATTCCTTAAGCTGAACGA 1088

QY 217 ggggatgaggttggtgacacaaatg 241
DB 1089 GGGGATCAAAATGCTGTTAATCAATG 1113

RESULT 15
LOCUS HUMCAP 3984 bp mRNA PRI 27-APR-1993
DEFINITION Human protein tyrosine phosphatase mRNA, complete cds.
VERSION M64572 M64572.1 GI:179912
KEYWORDS tyrosine phosphatase.
SOURCE Homo sapiens CDNA to mRNA.
ORGANISM Homo sapiens
DEFINITION Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

Search completed: January 31, 2002, 19:46:38
Job time: 1816 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 31, 2002, 19:16:11 ; Search time 1929.4 Seconds
(without alignments)
2238.936 Million cell updates/sec

Title: US-09-641-831-5
Perfect score: 402
Sequence: 1 atgatggaccactgaaga.....atcatgcactgcagcttga 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estom:*
4: em_estpl:*
5: em_estbl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estl:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	151.4	37.7	649	11	BG563384 602582446
2	100.8	25.1	1121	12	AK018481 Mus muscu
3	91.2	22.7	321	10	BB504666 BB504666
4	85.8	21.3	481	11	BF463332 UI-M-CGOp
5	84.8	21.1	484	10	AU170966 AU170966
6	82.6	20.5	231	10	AV368091 AV368091
7	76.4	19.0	597	11	BF037621 BF037621
8	72.2	18.0	558	11	BF434691 BF434691
9	64.6	16.1	383	13	AQ244687 HS_2045_A
10	63.8	15.9	354	10	AW001038 AW001038
11	60.2	15.0	309	13	AQ393528 CITBI-E1
12	60.2	15.0	543	13	AQ457616 HS_5097_B

C	13	60.2	15.0	560	13	AQ886165
C	14	59.2	14.7	875	11	BG432009
C	15	58.8	14.6	463	10	AI687027
C	16	57.6	14.3	372	10	BE763418
C	17	57.4	14.3	941	11	BG284962
C	18	57.2	14.2	265	10	AW166706
C	19	57.2	14.2	270	11	BG117722
C	20	57.2	14.2	325	10	AI885995
C	21	57.2	14.2	723	11	BF572167
C	22	57.2	14.2	540	11	AQ769741
C	23	56.8	14.1	404	11	R22776
C	24	56.8	13.9	860	13	AQ743955
C	25	55.8	13.9	824	10	AL042013
C	26	55.6	13.8	527	13	AQ313163
C	27	55.4	13.8	417	10	AI248313
C	28	55.4	13.8	457	10	AI248328
C	29	55.4	13.8	460	10	AI023290
C	30	55.4	13.8	495	11	BG231844
C	31	55.4	13.8	527	13	AQ494912
C	32	55.4	13.8	615	13	AQ768789
C	33	54.6	13.6	596	13	AQ569045
C	34	54.4	13.5	435	10	AI952612
C	35	54.2	13.5	441	13	AQ581530
C	36	54.2	13.5	499	13	AQ581533
C	37	54	13.4	389	10	AA748026
C	38	54	13.4	679	10	AA715049
C	39	54	13.4	932	13	CNS03SP4
C	40	53.8	13.4	214	13	AQ027606
C	41	53.8	13.4	307	13	AQ217481
C	42	53.8	13.4	513	13	AQ206820
C	43	53.8	13.4	583	13	AQ726346
C	44	53.8	13.4	593	10	BE153075
C	45	53.8	13.4	646	13	AQ583051

ALIGNMENTS

RESULT 1
BG563384 649 bp mRNA EST 10-APR-2001
LOCUS 602582446F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710299 5',
DEFINITION mRNA sequence.
ACCESSION BG563384.1 GI:13571036
VERSION BG563384.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 649)
AUTHORS NIH-MGC http://mgc.bcl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Title: Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM1548 row: 1 column: 12
High quality sequence stop: 548.
Location/Qualifiers
1. 649
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4710299"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1."

FEATURES

source
Location/Qualifiers
1. 649
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4710299"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1."


```
/clone = 050440004  
/clone_lib="RIKEN full-length enriched, 16 days embry
```

High quality sequence stop: 511.

FEATURES
 source
 1. 597
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3478523"
 /clone_lib="NCI_CGAP_Ov18"
 /tissue_type="fibrotheoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1; Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5', TGTTACCACTGAAGTGGAGCGCGCGACATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. *
 125 a 205 c 158 g 108 t 1 others

BASE COUNT
 ORIGIN

Query Match 19.0%; Score 76.4; DB 11; Length 597;
 Best Local Similarity 62.2%; Pred. NO. 1.9e-10;
 Matches 138; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

QY 131 gtggctgagcagcagagaccattaatcatctcaaggtcgaaagaaggggcaagcag 190
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 GCGGCGCGAGCAGCGCGAGCGCGTGGTCAATCACCAGATGAAGAGGGCGAGTAAAGC-- 58
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 191 acaccctgagctccaaactgcaggtgaggatgaggtgtgtgcacatcaatgaggtgactc 250
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 59 -CGCGCGGTGCACAGATTACTGGCTGGAGATGAGATCGTCGCGCATCAATGACATTGGTC 117
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 251 tgagcgcgtccagaagaaggaggcaggttccctggtgaaagatccacaagaccctcagc 310
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 118 TCTCAGGGTTTAGACAGAGGAGCGATTGGCTGGTGAAGGGGTCCTCAAGACCCCTGAAGC 177
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 311 tgrtagtcgcagaaaaatggggtcttctgtatgttgcccaaat 352
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 178 TGGTGTCAAAAGGAGGCGAGCTGGGCTGGAGGCCCTCACT 219
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8

BF434691
 LOCUS
 DEFINITION
 BP434691 558 bp mRNA EST 29-NOV-2000
 7p03b06.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644507 3'
 similar to SW:APXL_HUMAN Q13796 APICAL-LIKE PROTEIN ;, mRNA
 sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 558)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: ccapsb-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
 Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone Distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: -40UP from Glibco
 High quality sequence stop: 488.

FEATURES
SOURCE

Location/Qualifiers
1. .558

BASE COUNT
ORIGIN

a	191 c	147 g	102 t
---	-------	-------	-------

Query Match	18.0%;	Score 72.2;	DB 11;	Length 558;
Best Local Similarity	62.0%;	Pred. No. 2.5e-09;		
Matches 132;	Conservative	0;	Mismatches 78;	Indels 3;
				Gaps 1.

QY 140 agaacgggaacacatlaatcctccttaagtcggaagaaggggccaagcagacacctga 195
DB 1 AGCACGGCGAGCGCGTGTATCATCACCAAGATTGAAGGGCAGTAAAGC ---CGCGGCG 57
QY 200 gctccaacctcgaagctcyggagataggttctgcacatcaatgagctctgaacgct 259
DB 58 TCGAACAATCTACTGGCTTGAGATGAGATCGCGGCATCAATGACTTGCTCTCAAGGT 117
QY 260 ccagaagaaggagcagcttcctcctgttgaaagatctctcaagaacccctgaagctgtatgtc 319
DB 118 TTAGCACAGAAACCGATTGGCTGGTGGAAAGGGGTCCCTAAAGACCTGATGGTCTCA 177
QY 320 gcagaatcggggtctgtcatgttggccagat 352
DB 178 AAAGAGGAGACAGCTGGGCTGGAGGCTCTACT 210

[illegible]

FEATURES
source

```
Location/Qualifiers
1. .383
```

BASE COUNT
ORIGIN

86 a	92 c	115 g	90 f
------	------	-------	------

Query Match	16.1%;	Score 64.6;	DB 13;	Length 383;
Best Local Similarity	65.7%;	Pred. No. 2.6e-07;		
Matches 94;	Conservative 0;	Mismatches 49;	Indels 0;	Gaps 0;

[illegible]

RESULT	10			
AM001038/c				
LOCUS				
DEFINITION				
AM001038	354 bp	mrna	EST	08-MAR-2000
W711h03.x1	NCI CGAP	Rkid1	homo sapiens	CNA clone IMAGE:2495093 3
similar to	SM:APXL_HUMAN	Q13796	APICAL-LIKE	PROTEIN ; mRNA
sequence.				
AM001038				
AM001038.1	GI:5847954			
KEYWORDS				
SOURCE	human.			
ORGANISM	homo sapiens			

FEATURES	source	Location/Qualifiers
1..354		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone IMAGE:2495093"		
/clone_11b="NCI-CGAP_Kid11"		
/lab_host="DH10B"		
/note="Organ: kidney; Vector: pFRT3D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP_Kid3 was prepared, and ss circles were made in vitro. Following WAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs		

from a pool of 9,000 clones made from the same library (clonids 132376-132391, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

```

BASE COUNT      82 a      85 c      80 g      107 t
ORIGIN

Query Match      15.9%; Score 63.8; DB 10; Length 354;
Best Local Similarity 62.3%; Pred. No. 4.2e-07;
Matches 134; Conservative 0; Mismatches 77; Indels 4; Gaps 2;

Qy 111 ctggggtttacttaaggggtcctgggagcaggggaaccattatca-tctctaaag 169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 354 CTGGGGCTTACCCCTTAAGGGGCTCTGGAACCCCTGTGAACCCGCTCCCAAGTAAGA 295
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 170 tgaagaaggggaaagacagacccctgagctccaaactgcagctgggagatggtg 229
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 TTGAAGATGGAGGCAAGGCAG--CTTGTCCCAAGATGAGGACTGGTGATGAGCTGG 238
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 230 tgcacatcaatgagtgactctgagcagctccagaaaggagcaggttccctgggtgaag 289
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 237 TGAATATCAATGGCACTCCATATATGCTCCGCGCAAGAGGCGCTCATTTCTCATCAAG 178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 290 gatctacaagaccctcaggtcgttagtcgcaga 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 177 GCTCCTCGGATTCTCAAGCTGATTTGTCAGGAGA 143
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AQ393528/c 309 bp DNA GSS 06-MAR-1999
LOCUS CITBI-El-2553H18.TR CITBI-El Homo sapiens genomic clone 2553H18,
DEFINITION DNA sequence.
ACCESSION AQ393528
VERSION AQ393528.1 GI:4364551
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 309)
REFERENCE Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
AUTHORS Venter,J.C.
TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: CITBI-El-2553H18.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: bhe@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
source 1..309
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2553H18"
/clone_lib="CITBI-El"
/sex="male"
/cell_type="sperm"
/note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT 96 a 48 c 70 g 94 t 1 others
ORIGIN

Query Match      15.0%; Score 60.2; DB 13; Length 543;
Best Local Similarity 84.0%; Pred. No. 3.9e-06;
Matches 68; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 322 agaaatgggggtctgtctatgttggcccaagaatggaaggtagtggtctattcatagcgat 381
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 ACAGATGAAGTCTTGTCTATGTTGCCAGGCTGGCAGTGGTGTATTTCACAGGCGATGAT 44
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 382 catcatgcaactgcagccttga 402
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43 CATCATGGACACAGCGCTTGA 23
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AQ457616/c 543 bp DNA GSS 23-APR-1999
LOCUS HS_5097_B1_H10_T7A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate-673 Col-19 Row-P, DNA sequence.
ACCESSION AQ457616
VERSION AQ457616.1 GI:4636256
KEYWORDS GSS.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 543)
REFERENCE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
AUTHORS Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 93380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@jeng.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 673 row: P column: 19
Seq primer: T7
Class: BAC ends
High quality sequence stop: 543.
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate-673 Col-19 Row-P"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 206 a 94 c 75 g 159 t 9 others
ORIGIN

Query Match      15.0%; Score 60.2; DB 13; Length 543;
Best Local Similarity 84.0%; Pred. No. 4.4e-06;
Matches 68; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 322 agaaatgggggtctgtctatgttggcccaagaatggaaggtagtggtctattcatagcgat 381
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ORIGIN

```

ACCESSION	AI687027
VERSION	AI687027.1
	GI:4898321

Fri Feb 1 08:56:34 2002

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 463)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1288 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 418.
 FEATURES Location/Qualifiers
 1..463
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2206733"
 /clone_lib="NCI-CGAP-Ut3"
 /tissue_type="poorly-differentiated endometrial
 adenocarcinoma, 2 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
 Average insert size 1.45 kb. Life Technologies catalog #:
 11541-018"
 BASE COUNT 144 a 105 c 87 g 125 t 2 others
 ORIGIN
 Query Match 14.6%; Score 58.8; DB 10; Length 463;
 Best Local Similarity 84.6%; Pred. No. 1e-05;
 Matches 66; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 322 acaaatgggggtttgtatgttgcacgaatgaaggtagtggtatttcacagcagatg 381
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2 AGAGACAGGGTCTTGCTATGTGCCAGGCTGAGTGTAGTGGCTATTTCACAGCAGAT 61
 QY 382 catcagcactgcagcct 399
 ||| ||||| |||||
 Db 62 CATAGTCACAGCCTT 79
 Search completed: January 31, 2002, 19:16:15
 Job time: 1963 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:46:38 ; Search time 1599.5 Seconds
(without alignments)
4146.212 Million cell updates/sec

Title: US-09-641-831-5

Perfect score: 402
Sequence: 1 atgatgagcaccactgaaga.....atcatgcactgcagccttga 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenBank:
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_ov:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hgtgo_hum:*
31: em_hgtgo_inv:*
32: em_hgtgo_rod:*
33: em_hgtgo_hum:*
34: em_hgtgo_inv:*
35: em_hgtgo_rod:*
36: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	402	100.0	402	6	AX088027 Sequence
2	323	80.3	327	6	AX088025 Sequence
3	323	80.3	573	6	AX088023 Sequence
4	211.2	52.5	6423	10	AF199421 Mus muscu
5	170	42.3	163535	2	AC046189 Homo sapi
6	158.6	39.5	156375	2	AP002859
7	98.8	24.6	7445	9	HSAPXL
8	85.2	21.2	6014	9	AB033028
9	81	20.1	156375	2	AP002859
10	64.6	16.1	196542	9	AL512843
11	63.6	15.8	75684	2	AL1390030
12	62.8	15.6	87451	2	AL1390030
13	62	15.4	127639	2	HSBJ79963
14	62	15.4	168957	2	AC025079
15	61.8	15.4	135693	2	HSJ116744
16	61.6	15.3	160770	2	AC016796
17	61	15.2	113664	9	AC020719
18	61	15.2	113664	9	HS12803
19	61	15.2	151517	2	AC031989
20	60.6	15.1	193126	2	CNS018P3
21	60.2	15.0	45262	9	AL391689
22	60.2	15.0	46534	9	HS1057D18
23	60.2	15.0	106687	9	AL356124
24	60.2	15.0	123331	2	AC010942
25	60.2	15.0	140730	2	AC011886
26	60.2	15.0	157103	2	AC004958
27	60.2	15.0	163031	2	AC024947
28	60.2	15.0	163889	2	AL451010
29	60.2	15.0	183480	2	AL390040
30	60.2	15.0	186350	2	AC074325
31	60.2	15.0	202304	2	AC079240
32	60.2	15.0	202924	2	AC016397
33	60.2	15.0	206903	9	AL136971
34	59.4	14.8	82601	2	HSJ344H20
35	59.4	14.8	142441	2	AC018545
36	59.4	14.8	170242	2	AC079121
37	59.2	14.7	190030	2	CNS00007
38	59.2	14.7	195881	9	AC007055
39	59.2	14.7	199927	9	AL365505
40	59	14.7	101500	9	AC013399
41	59	14.7	171812	2	AC012447
42	58.8	14.6	42213	9	AC010527
43	58.8	14.6	42213	9	AC092409
44	58.8	14.6	156294	2	AC067984
45	58.8	14.6	157531	2	AC067984

ALIGNMENTS

RESULT 1
AX088027
LOCUS AX088027
DEFINITION Sequence 5 from Patent WO0114422.
ACCESSION AX088027
VERSION AX088027.1 GI:13396953
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 402)
Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.

AUTHORS
TITLE Sequence derived from a human mammary gland cdna library
Patent: WO 0114422-A 5 01-MAR-2001;
JOURNAL Lexicon Genetics Incorporated (US)
Location/Qualifiers

FEATURES
source 1..402
/organism="Homo sapiens"
/db_xref="taxon:9606"


```
BASE COUNT      110 a      91 c      117 g      84 t
ORIGIN

Query Match      100.0%; Score 402; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.2e-110;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgatgagacactgaagacttcacaaagcctagtgccacattaaactctaaacacggcc 60
Db 1 ATGATGAGGACCACTGAAGACTTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 60
Qy 61 accaagggaaggtacatttattctgagggcattctgagggagagctccctggggtttt 120
Db 61 ACCAAGGGAAGGTACATTATCTGAGGCACTCTCTGAGGGAGGAGCTCCCTGGGGTTTT 120
Qy 121 actctaaagggtggcctggagcagcgagaaaccattatctctctaaagtggaagaagg 180
Db 121 ACTCTAAAGGCTGGCTGGAGCAGCGAGAACCATTAATCATCTCTTAAGGTGGAAGAAGG 180
Qy 181 ggcacagacacacccctgagctccaaactgagctgggagtgaggttggcacatcaat 240
Db 181 GGCACAGACACACCCCTGAGCTCCAAACTGAGCTGGGATGAGGTGTGTGCACATCAAT 240
Qy 241 gaggtgactctgagcagctccagaaaggagcagtttccctgggtgaaaggatctctacaag 300
Db 241 GAGGTGACTCTGAGCAGCTCCAGAAAGGAGCGAGTTTCCCTGGTGTGAAAGGATCCTACAAG 300
Qy 301 accctcaggtggtagtcgcag 323
Db 301 ACCCTCAGGCTGGTAGTGGCGAG 323

RESULT 3
AX088023 LOCUS AX088023 573 bp DNA PAT 17-MAR-2001
DEFINITION Sequence 1 from Patent WO0114422.
ACCESSION AX088023
VERSION AX088023.1 GI:13396951
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 573)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 1 01-MAR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Source 1..573
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 151 a 152 c 152 g 118 t
ORIGIN

Query Match 80.3%; Score 323; DB 6; Length 573;
Best Local Similarity 100.0%; Pred. No. 1.4e-86;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgatgagacactgaagacttcacaaagcctagtgccacattaaactctaaacacggcc 60
Db 1 ATGATGAGGACCACTGAAGACTTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 60
Qy 61 accaagggaaggtacatttattctgagggcattctgagggagagctccctggggtttt 120
Db 61 ACCAAGGGAAGGTACATTATCTGAGGCACTCTCTGAGGGAGGAGCTCCCTGGGGTTTT 120
Qy 121 actctaaagggtggcctggagcagcgagaaaccattatctctctaaagtggaagaagg 180
Db 121 ACTCTAAAGGCTGGCTGGAGCAGCGAGAACCATTAATCATCTCTTAAGGTGGAAGAAGG 180
Qy 181 ggcacagacacacccctgagctccaaactgagctgggagtgaggttggcacatcaat 240
Db 181 GGCACAGACACACCCCTGAGCTCCAAACTGAGCTGGGATGAGGTGTGTGCACATCAAT 240
Qy 241 gaggtgactctgagcagctccagaaaggagcagtttccctgggtgaaaggatctctacaag 300
Db 241 GAGGTGACTCTGAGCAGCTCCAGAAAGGAGCGAGTTTCCCTGGTGTGAAAGGATCCTACAAG 300
Qy 301 accctcaggtggtagtcgcag 323
Db 301 ACCCTCAGGCTGGTAGTGGCGAG 323

RESULT 4
AF199421 LOCUS AF199421 6423 bp mRNA ROD 13-DEC-1999
```

```
BASE COUNT      110 a      91 c      117 g      84 t
ORIGIN

Query Match      100.0%; Score 402; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.2e-110;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgatgagacactgaagacttcacaaagcctagtgccacattaaactctaaacacggcc 60
Db 1 ATGATGAGGACCACTGAAGACTTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 60
Qy 61 accaagggaaggtacatttattctgagggcattctgagggagagctccctggggtttt 120
Db 61 ACCAAGGGAAGGTACATTATCTGAGGCACTCTCTGAGGGAGGAGCTCCCTGGGGTTTT 120
Qy 121 actctaaagggtggcctggagcagcgagaaaccattatctctctaaagtggaagaagg 180
Db 121 ACTCTAAAGGCTGGCTGGAGCAGCGAGAACCATTAATCATCTCTTAAGGTGGAAGAAGG 180
Qy 181 ggcacagacacacccctgagctccaaactgagctgggagtgaggttggcacatcaat 240
Db 181 GGCACAGACACACCCCTGAGCTCCAAACTGAGCTGGGATGAGGTGTGTGCACATCAAT 240
Qy 241 gaggtgactctgagcagctccagaaaggagcagtttccctgggtgaaaggatctctacaag 300
Db 241 GAGGTGACTCTGAGCAGCTCCAGAAAGGAGCGAGTTTCCCTGGTGTGAAAGGATCCTACAAG 300
Qy 301 accctcaggtggtagtcgcagaaatggggtcttgctatgttggccagaatggaagta 360
Db 301 ACCCTCAGGCTGGTAGTGCGAGAAATGGGTCTTGCTATGTGCCAGAAATGGAAGTA 360
Qy 361 gtggctattcatgagcatgatcatcatcagctcagccttga 402
Db 361 GTGGCTATTTCATAGGCATGATCATCATGCTGAGCCTTGA 402

RESULT 2
AX088025 LOCUS AX088025 327 bp DNA PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114422.
ACCESSION AX088025
VERSION AX088025.1 GI:13396952
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 327)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 3 01-MAR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Source 1..327
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 91 a 76 c 96 g 64 t
ORIGIN

Query Match 80.3%; Score 323; DB 6; Length 327;
Best Local Similarity 100.0%; Pred. No. 1.4e-86;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgatgagacactgaagacttcacaaagcctagtgccacattaaactctaaacacggcc 60
Db 1 ATGATGAGGACCACTGAAGACTTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 60
Qy 61 accaagggaaggtacatttattctgagggcattctgagggagagctccctggggtttt 120
Db 61 ACCAAGGGAAGGTACATTATCTGAGGCACTCTCTGAGGGAGGAGCTCCCTGGGGTTTT 120
```

DEFINITION Mus musculus PDZ domain actin binding protein Shroom mRNA, complete cds.

ACCESSION AF199421

VERSION AF199421.1

KEYWORDS GI:6467989

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 6423)
Hildebrand, J.D. and Soriano, P.
Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice
Cell 99 (5), 485-497 (1999)

JOURNAL 20055594

MEDLINE 10589677

PUBMED 10589677

REFERENCE 2 (bases 1 to 6423)
Hildebrand, J.D. and Soriano, P.
Direct Submission
Submitted (27-OCT-1999) Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA

FEATURES

source Location/Qualifiers

1..6423

/organism="Mus musculus"

/db_xref="taxon:10090"

233..6193

/note="shrm"

/codon_start=1

/product="PDZ domain actin binding protein Shroom"

/protein_id="AF13269.1"

/db_xref="GI:6467989"

/translation="MKTPELEPSATPNSPTPTERTFYLLALLLEGAPWGTLLKGG
LEGEPLITISKEGKADSVSSGLQAGDEVHININVALSPREAVSLVKGSTKRL
LYRRVVCAGAPGADPGTSLKSLSSGLTSPGKATWSGVALIKORCESEATPH
SWITTFEGTQPVSMQISQGMPPMHWSSSTSLSYDHAFLRRSPDOCS
QGMESLEPSGTPCHLSPAKSTISIDQLGHLNKRDAISFSTSSIFEPYPPG
GSARERGSMDVISARGGLLEGKQADIRVKTIVTTRKVSSEYEVKAPLLQGRD
AHASADSGCAKMSIIPRGKTPSPSQSCSLFADNDLPQKAGAPLPPRSQSY
AARFRERPSMSLDRKRCRQPTNSGOKTFFEDQHLTVPERSPNSPVSKH
NYOKAKOPGLPTGIYVSPDPHPAQAQVOPSVSNGVYALVYESGYTAAGTC
NKMATIDENONENASPRGAPCOLEHNSVTPEKREPTAKYIVYVHSVEVND
SSLRHITTPHGHSPPTSEKNTHGSRACSNHSLSSPOAKLHVGDNRPRLSQP
WEDFOEDHNAHROKVERGQGLSGNSGKTRSAFSLQNTPELRSONVELGSA
QEVHPGSRKVEDPGRKAGASDIRGLDRSVTPPRGKNADVSAVSASRIBESPA
PALPOTSGASORLSSSSAAPOYRKHPCVLEKVSRIEERDGRHRLPGSSAYCP
GNRGRTGPTSSSDLDPKAGSVHSETEHLRNGENQNGEAKOEASPPQCS
HLIRARADGRPPARGGPRPARLLRSOSTOLYSEAREASVEDRPPRESPL
LDAPFSRATKNSIDKQSRVLCATSEFRRLDLEPCTPATSPRPPRPAHVGMSPPA
AVPSSPHTPRERHSVTPAQAARPRKRLRYEOKKRSYSEPKNVEGVSEAPR
TPCGPPRPAQPRESESTVADRRLFEERDGRACSTLSGPELKOQNALADYQRT
GKRTGASHTGGAARARATERLPPGRPRGAKWPRLASCLSLSRLEPRLKREH
PSADGPOAPDRSSSPASGRVLERRRMDPOVROLISGANCEPVRORDGAGGP
PSMGVAGKAKASADLERSDTLAVPYVRKSSPTSDKQODVLEHSGNEGY
KDPCCLAGPGRSLSCDKQNETALPLHHTPCWNGSGKATVASSAPSSGAAH
LKORARPRPLSKMGHPRDAAALSLSLSPSPASPVSPVSSQALAMDQOTG
QPPSPASAVIQTSPSLSSSLSPAYGLGEMKRTSLPQRPWPWYKMAVVEDCL
AEDTLAPFAFKLHYRNPSPSSSDPTPERISIRISLQSGAPPKCYDDE
VFMKDLHKVTSPTFEALPPPPSPSEPLVNGTDPPPPPPQALCVLLDGEA
STEGSGPCRIPRVMTREGHVPGAASEGQIMATIPQPSAGSEASESPSSASA
QPOLNSPGKOLQSONRLTYEVEVERODLCKTHAEPQTSKDIRLEALAKEIVH
DKSLADIDPPSRMKTMDMEGLPGDAVILMSGAKRLADITARRAGCEARSH
KEASVLNCFATYSVSAKAKELINKIKDMEPELOEOPGQVYKKAELSLHLKL
ESLQKAVSLTLIDIKLNALDEEVALISELCKRNEPKYIMFGIDDKVYNNLLSL
GRLAREVNLGLGEDAKESERSLNEKRYLAQOHEPARLAKEDREVDLYDILA
NYLSEALQDYQHFYKMKSTLLIEDRKLDXIKLQEQEVRLCLSLPSDFPRKGAIS
LPPALTGATGTSVFGVPEPTLTSPL"

BASE COUNT 1422 a 2042 c 1853 g 1099 t 7 others

ORIGIN

Query Match 52.5%; Score 211.2; DB 10; Length 6423;

Best Local Similarity 77.7%; Pred. No. 8e-53;

Matches 255; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 3 gatgagaccactgaagactccacaagcctagtcacatcaatcaacgagccac 62

Db 232 GATGAAGACTCCGAGAACTTGAGAGAGCCGACAGCCAAACCCAGTAGAGACCC 291

QY 63 caaggaaggtacattatctctggaagcattcctggaaggaaggtccctggtttac 122

Db 292 CACGAGAGACTTCGTTTACTGGAAGCCGCTCTAAGGAGGAGGCTCCCTGGGCTTAC 351

QY 123 tctaaaggtgctccggagcacgaagacatcaatcattcctggaaggaaggg 182

Db 352 CTTGAAGTGGCTTGAGACGCTGAGAACCGTTATCATTTCCAGATGGAAGAGGGG 411

QY 183 caagcacagaccctgagctcacaactgcagctgtggatgaggtgtgacataatga 242

Db 412 CAAAGCCACTGAGTACGCTCCGAGCTGCGGAGGAGTCAATACATCAATGA 471

QY 243 ggtgactctgagcagctcagaagaagagcagttccctggtgaagagatcctacaagc 302

Db 472 GGTGCGCTGAGCAGTCCAGAGAGAGGCTCTCTCCCTAGTGAAGGCTCTACAGAC 531

QY 303 cctcagctgtagtgagcagaatgg 330

Db 532 CCTGCGGCTGCTGCTGCTGCGAGACGCTG 559

RESULT 5

AC046189

LOCUS

DEFINITION Homo sapiens chromosome 4 clone RP11-356M17 map 4, WORKING DRAFT

ACCESSION AC046189

VERSION AC046189.4

KEYWORDS GI:10047902

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 163535)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 4, clone RP11-356M17
Unpublished

2 (bases 1 to 163535)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barrow, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkert, G., Campilioni, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laboche, K., Lamazares, R., Landers, T., Lehocsky, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McNeely, R., Meldrum, T., Meneses, L., Minova, T., Miranda, C., Meng, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Sudramanian, A., Talamas, J., Testa, S., Theodore, J., Tittel, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (13-ARR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 10, 2000 this sequence version replaced gi:8083636.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIGR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9512

Center clone name: 356_M.17

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 148971 bases at least Q40

Consensus quality: 156340 bases at least Q30

Consensus quality: 159016 bases at least Q20

Insert size: 181000; agarose-fp

Insert size: 160935; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 2095: contig of 2095 bp in length
* 2096 2195: gap of 100 bp
* 2196 3520: contig of 1325 bp in length
* 3521 3620: gap of 100 bp
* 3621 4777: contig of 1157 bp in length
* 4778 4877: gap of 100 bp
* 4878 6660: contig of 1783 bp in length
* 6661 6760: gap of 100 bp
* 6761 8530: contig of 1770 bp in length
* 8531 8630: gap of 100 bp
* 8631 11149: contig of 2519 bp in length
* 11150 11249: gap of 100 bp
* 11250 15298: contig of 4049 bp in length
* 15299 15398: gap of 100 bp
* 15399 18533: contig of 3135 bp in length
* 18534 18633: gap of 100 bp
* 18634 22044: contig of 3411 bp in length
* 22045 22144: gap of 100 bp
* 22145 25578: contig of 3434 bp in length
* 25579 25678: gap of 100 bp
* 25679 29571: contig of 3893 bp in length
* 29572 29671: gap of 100 bp
* 29672 34405: contig of 4734 bp in length
* 34406 34505: gap of 100 bp
* 34506 37415: contig of 2910 bp in length
* 37416 37515: gap of 100 bp
* 37516 44162: contig of 6647 bp in length
* 44163 44262: gap of 100 bp
* 44263 48837: contig of 4575 bp in length
* 48838 48937: gap of 100 bp
* 48938 55115: contig of 6178 bp in length
* 55116 55215: gap of 100 bp
* 55216 62194: contig of 6979 bp in length
* 62195 62294: gap of 100 bp
* 62295 69547: contig of 7253 bp in length
* 69548 69647: gap of 100 bp
* 69648 78050: contig of 8403 bp in length
* 78051 78150: gap of 100 bp
* 78151 88041: contig of 9891 bp in length
* 88042 88141: gap of 100 bp
* 88142 98560: contig of 10419 bp in length
* 98561 98660: gap of 100 bp
* 98661 108146: contig of 9486 bp in length
* 108147 108246: gap of 100 bp
* 108247 118855: contig of 10609 bp in length
* 118856 118955: gap of 100 bp

* 118956 128543: contig of 9588 bp in length
* 128544 128643: gap of 100 bp
* 128644 142647: contig of 14004 bp in length
* 142648 142747: gap of 100 bp
* 142748 153569: contig of 10822 bp in length
* 153570 153669: gap of 100 bp
* 153670 163535: contig of 9866 bp in length.

FEATURES

source

1..163535

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-356M17"

/clone_lib="RPC1-11 Human Male BAC"

1..2095

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

2196..3520

/note="assembly_fragment"

3621..4777

/note="assembly_fragment"

4878..6660

/note="assembly_fragment"

6761..8530

/note="assembly_fragment"

8631..11149

/note="assembly_fragment"

11250..15298

/note="assembly_fragment"

15399..18533

/note="assembly_fragment"

18634..22044

/note="assembly_fragment"

22145..25578

/note="assembly_fragment"

25679..29571

/note="assembly_fragment"

29672..34405

/note="assembly_fragment"

34506..37415

/note="assembly_fragment"

37516..44162

/note="assembly_fragment"

44263..48837

/note="assembly_fragment"

48938..55115

/note="assembly_fragment"

55216..62194

/note="assembly_fragment"

62295..69547

/note="assembly_fragment"

69648..78050

/note="assembly_fragment"

78151..88041

/note="assembly_fragment"

88142..98560

/note="assembly_fragment"

98661..108146

/note="assembly_fragment"

108247..118855

/note="assembly_fragment"

118956..128543

/note="assembly_fragment"

128644..142647

/note="assembly_fragment"

142748..153569

/note="assembly_fragment"

153670..163535

/note="assembly_fragment"

clone_end:T7

vector_side:right"

BASE COUNT	48438 a	32205 c	31514 g	48770 t	2608 others
Query Match					
Best Local Similarity	100.0%;	Pred. No. 2,5e-40;			
Matches 170; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 1	atgataagaccactgaagacttccaaagctagtgcacataaactcaacagcgc 60				
Db 62580	ATGATGAGAGACCACTGAGAGACTTCCACAGAGCTATGTCACATTAACTCTAACAGCGCC 62639				
QY 61	accaaggaaggtacattatcctgaagcattcctctgtagggagagagcctccctgggtttt 120				
Db 62640	ACCAAGGAGAGTACATTATTACTGGAGGCATTCCTGAGGAGGAGAGCTCCCTGGGGTWT 62699				
QY 121	actctaaagggtgacctgagacgaggaacacattatcatctcctaagt 170				
Db 62700	ACTCTAAGAGGCTGCGCTGAGACGAGAGACCATTAATCATCTTAAGT 62749				
RESULT 6					
AP002859	156375 bp	DNA	HTG	29-AUG-2000	
LOCUS	Homo sapiens Chromosome 4 clone 225E15	map 4q16-q20, *** SEQUENCING			
DEFINITION	IN PROGRESS ***, 15 unordered pieces.				
ACCESSION	AP002859				
VERSION	AP002859.1	GI:9955382			
KEYWORDS	HTG; HTGS, PHASE1.				
SOURCE	Homo sapiens DNA, clone:225E15.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 156375)				
JOURNAL	Tsai,S.F.				
COMMENT	Direct Submission Submitted (28-AUG-2000) Shih-Feng Tsai, National Yang-Ming University, Institute of Genetics, 155 Li-Rong St. Section 2, Beitou, Taipei, Taiwan 11221, Republic of China (E-mail:ympeisai@ym.edu.tw, URL:http://genome.ym.edu.tw//, Tel:886-2-28267043, Fax:886-2-28264930) These sequences are draft human sequences, not finished sequences. These sequences are unordered pieces. And gaps between the contigs of the same clone are represented as 100 N. * NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will be preserved. 1 21185: contig of 21185 bp in length * 21186 21285: gap of 100 bp * 21286 22928: contig of 1643 bp in length * 22929 23028: gap of 100 bp * 23029 26285: contig of 3257 bp in length * 26286 26385: gap of 100 bp * 26386 30606: contig of 4221 bp in length * 30607 30706: gap of 100 bp * 30707 36924: contig of 6218 bp in length * 36925 37024: gap of 100 bp * 37025 40953: contig of 3929 bp in length * 40954 41053: gap of 100 bp * 41054 47843: contig of 6790 bp in length * 47844 47943: gap of 100 bp * 47944 57174: contig of 9231 bp in length * 57175 57274: gap of 100 bp * 57275 64994: contig of 7720 bp in length * 64995 65094: gap of 100 bp * 65095 76523: contig of 11429 bp in length * 76524 76623: gap of 100 bp * 76624 88306: contig of 11683 bp in length				

[illegible]


```

QY 214 gctgggagtgagtggtgtgacataatgagtgacttcgacgctccagaaggagca 273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 186 ACTGGGAGTACCTGGTGAATATCATGATGGCACTCATTTATAGGCTCCGCCAAGAGGCC 245
QY 274 gttccctgtgaagaagacaccacagcctcagctgtgtgtgcgcagaatgggtc 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 246 CTCATCTCATCATCAAGAGCCCTTCGCCGATTTCTCAAGCTGATTCAGAGAGCAAGCC 305
QY 334 ttgctatgttgcaccagaatgagtgagtgcta 367
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 306 CCTGTCAGTAGGCGCCACATCATGCGATGTGGCCA 339

RESULT 9
AP002859/c 156375 bp DNA HTG 29-AUG-2000
LOCUS Homo sapiens chromosome 4 clone 225E15 map 4q16-q20, *** SEQUENCING
DEFINITION IN PROGRESS ***, 15 unordered pieces.
ACCESSION AP002859
VERSION AP002859.1 GI:9955382
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens DNA, clone:225E15.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 156375)
REFERENCE Tsai,S.F.
          Submitted (28-AUG-2000) Shih-Feng Tsai, National Yang-Ming
          University, Institute of Genetics, 155 Li-Rong St. Section 2,
          Peitou, Taipei, Taiwan 11221, Republic of China
          (E-mail:lympetsai@ym.edu.tw, URL:http://genome.ym.edu.tw/,
          Tel:866-2-28267043, Fax:866-2-28264930)
          These sequences are draft human sequences, not finished sequences.
          These sequences are unordered pieces. And gaps between the contigs
          of the same clone are represented as 100 N.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 15 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          1
          21185: contig of 21185 bp in length
          21186 21285: gap of 100 bp
          21286 22928: contig of 1643 bp in length
          22929 23028: gap of 100 bp
          23029 26285: contig of 3257 bp in length
          26286 26385: gap of 100 bp
          26386 30606: contig of 4221 bp in length
          30607 30706: gap of 100 bp
          30707 36924: contig of 6218 bp in length
          36925 37024: gap of 100 bp
          37025 40953: contig of 3929 bp in length
          40954 41053: gap of 100 bp
          41054 47843: contig of 6790 bp in length
          47844 47943: gap of 100 bp
          47944 57174: contig of 9231 bp in length
          57175 57274: gap of 100 bp
          57275 64994: contig of 7720 bp in length
          64995 65094: gap of 100 bp
          65095 76523: contig of 11429 bp in length
          76524 76623: gap of 100 bp
          76624 88306: contig of 11683 bp in length
          88307 88406: gap of 100 bp
          88407 101162: contig of 12756 bp in length
          101163 101262: gap of 100 bp
          101263 115436: contig of 14174 bp in length
          115437 115536: gap of 100 bp
          115537 133350: contig of 17814 bp in length
          133351 133450: gap of 100 bp

```

```

* 133451 156375: contig of 22925 bp in length.
  Location/Qualifiers
    source
      1. 156375
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="4"
        /map="4q16-q20"
        /clone="225E15"

BASE COUNT 45683 a 32392 c 31715 g 45174 t 1411 others
ORIGIN

Query Match 20.1%; Score 81; DB 2; Length 156375;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 agaattggggtctgtctatgttgcaccagaatgagtgagtgctattcatagcagatg 381
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132700 AGAAATGGGCTTCTCTATGTGTCGCCCAATGGAGTAGTGCTATTCATATGCGCATGAT 132641
QY 382 catcatgcacttcgacgttga 402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132640 CATCATGCACTGCGACGCTTGA 132620

RESULT 10
AC002365 196542 bp DNA PRI 22-AUG-1998
LOCUS Homo sapiens chromosome X clone U177G4, U152H5, U168D5, 17A46,
DEFINITION U172D6, and U186B3 from Xp22, complete sequence.
ACCESSION AC002365 U82630 U83511
VERSION AC002365.1 GI:2358015
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 196542)
REFERENCE Muzny,D., Arenson,A.D., Adams,C., Brundage,E., Bunac,C.,
          Curran,J., Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M.,
          Karpal,J., Karpal,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O.,
          Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswald,G., Perez,L.,
          Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,
          Simon,M., Stovall,K., Timms,K.M., Todd,J., Vo,Q., Williamson,A.,
          Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
          Direct Submission
          Unpublished
          2 (bases 1 to 196542)
REFERENCE Chiu,M.W.
          Direct Submission
          Submitted (23-JUL-1997) Molecular and Human Genetics, Baylor
          College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
          3 (bases 1 to 196542)
REFERENCE Muzny,D., Arenson,A.D., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
          Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Gonzalez,O.,
          Haywood,M., Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L.,
          Shen,H., Worley,K., Chen,E., Forcum,J., Arenson,A.D., Chiu,M.W.,
          Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and
          Gibbs,R.A.
          Direct Submission
          Submitted (20-AUG-1997) Molecular and Human Genetics, Baylor
          College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
          4 (bases 1 to 196542)
REFERENCE Chiu,M.W.
          Direct Submission
          Submitted (04-SEP-1997) Molecular and Human Genetics, Baylor
          College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA
          5 (bases 1 to 196542)
REFERENCE Worley,K.C.
          Direct Submission
          Submitted (22-AUG-1998) Human Genome Sequencing Center, Department

```


repeat_region	/note="L2 repeat: matches 1228. .2345 of consensus" 38554. .38703
repeat_region	/note="L2 repeat: matches 2074. .2225 of consensus" 41312. .41373
repeat_region	/note="MIR repeat: matches 104. .176 of consensus" 41372. .42513
repeat_region	/note="L2 repeat: matches 1610. .2745 of consensus" 42530. .42693
repeat_region	/note="MIR repeat: matches 60. .247 of consensus" 44660. .44787
repeat_region	/note="MIR repeat: matches 20. .160 of consensus" 45322. .45573
repeat_region	/note="MER46C repeat: matches 46. .328 of consensus" 46112. .46244
repeat_region	/note="L2 repeat: matches 2576. .2709 of consensus" 46264. .46406
repeat_region	/note="L2 repeat: matches 2411. .2558 of consensus" 46520. .46943
repeat_region	/note="L2 repeat: matches 1702. .2148 of consensus" 46954. .47188
repeat_region	/note="L2 repeat: matches 1951. .2215 of consensus" 47240. .47315
repeat_region	/note="L2 repeat: matches 1789. .1857 of consensus" 47316. .47619
repeat_region	/note="AluSg repeat: matches 1. .306 of consensus" 47620. .47908
repeat_region	/note="L2 repeat: matches 1461. .1789 of consensus" 47938. .49004
repeat_region	/note="MER11C repeat: matches 1. .1057 of consensus" 49018. .49086
repeat_region	/note="L2 repeat: matches 1389. .1462 of consensus" 49660. .49750
repeat_region	/note="MLT1J repeat: matches 103. .199 of consensus" 51571. .51786
repeat_region	/note="MIR repeat: matches 35. .247 of consensus" 51747. .51971
repeat_region	/note="MIR repeat: matches 7. .224 of consensus" 52048. .52353
repeat_region	/note="AluX repeat: matches 1. .307 of consensus" 52905. .53006
repeat_region	/note="51 copies 2 mer to 60% conserved" 53361. .53422
repeat_region	/note="31 copies 2 mer ca 71% conserved" 53471. .53579
repeat_region	/note="L2 repeat: matches 2628. .2736 of consensus" 53754. .53916
repeat_region	/note="L1MC/D repeat: matches 5375. .5396 of consensus" 53917. .54205
repeat_region	/note="AluSg repeat: matches 1. .289 of consensus" 54206. .54500
repeat_region	/note="L1MC/D repeat: matches 5094. .5375 of consensus" 55056. .55280
repeat_region	/note="MIR repeat: matches 3. .262 of consensus" 55391. .55644
repeat_region	/note="MIR repeat: matches 26. .261 of consensus" 56355. .56637
repeat_region	/note="AluX repeat: matches 21. .311 of consensus" 56650. .56786
repeat_region	/note="L1pA13 repeat: matches 5998. .6144 of consensus" 57041. .57355
repeat_region	/note="AluB8 repeat: matches 1. .315 of consensus" 57476. .57556
repeat_region	/note="L2 repeat: matches 778. .857 of consensus" 60400. .60703
repeat_region	/note="L2 repeat: matches 2404. .2750 of consensus" 60950. .61243
repeat_region	/note="AluTb repeat: matches 1. .294 of consensus" 61616. .61838
repeat_region	/note="MIR repeat: matches 5. .241 of consensus" 62351. .62581
repeat_region	/note="MIR repeat: matches 22. .262 of consensus" 62881. .63097
repeat_region	/note="MIR repeat: matches 18. .254 of consensus"

```

repeat_region 63440..63598
/note="MIR repeat: matches 102..252 of consensus"
repeat_region 63599..63840
/note="AluSq repeat: matches 67..307 of consensus"
repeat_region 64159..64563
/note="L2 repeat: matches 2273..2710 of consensus"
repeat_region 64868..64960
/note="MIR repeat: matches 90..200 of consensus"
repeat_region 65432..66100
/note="MIR repeat: matches 90..200 of consensus"

Query Match 15.8%; Score 63.6; DB 9; Length 75684;
Best Local Similarity 76.5%; Pred. No. 2.8e-08;
Matches 78; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 301 accctcagctggtagtcgcagaaatgggtctctgtatgttgcgcagaatggaaggta 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31278 AGCCACACTTATTATTATTAAGAGATGGGCTGTGTATGTTGCCAGACTGAAATGCA 31337
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 361 gtggctatctagcgatgatcatcatcgactgacgacctga 402
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31338 GTGGCTATTACAGGCCCTGATGATGACTCACTGCGAGCTTTGA 31379
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AL390030 87451 bp DNA HTG 20-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP4-700A9, *** SEQUENCING IN
DEFINITION PROGRESS ***, 9 unordered pieces.
ACCESSION AL390030
VERSION AL390030.2 GI:9368769
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Pavitt,R.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 22, 2000 this sequence version replaced gi:9368315.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj700A9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 10% of reads
Dye-terminator Big Dye; 89% of reads
Consensus quality: 82540 bases at least Q40
Consensus quality: 84030 bases at least Q30
Consensus quality: 85000 bases at least Q20
Insert size: 86651; sum-of-contigs
Insert size: 101143; 4.6% error; agarose-fp
Quality coverage: 4.16x in Q20 bases; sum-of-contigs Quality
coverage: 3.57x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* * 1 26100: contig of 26100 bp in length
* * 26101 26200: gap of 100 bp
* * 26201 33791: contig of 7591 bp in length

```

```

* 33792 33891: gap of 100 bp
* 33892 43982: contig of 10091 bp in length
* 43983 44082: gap of 100 bp
* 44083 55959: contig of 11877 bp in length
* 55960 56059: gap of 100 bp
* 56060 65631: contig of 9572 bp in length
* 65632 65731: gap of 100 bp
* 65732 68376: contig of 2645 bp in length
* 68377 68476: gap of 100 bp
* 68477 72701: contig of 4225 bp in length
* 72702 72801: gap of 100 bp
* 72802 78093: contig of 5292 bp in length
* 78094 78193: gap of 100 bp
* 78194 87451: contig of 9258 bp in length.
FEATURES
Source
1..87451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-790A9"
/clone_1bp="RPCI-4"
1..26100
/note="assembly-fragment:00202"
misc_feature
/clone_end:SP6
vector_side:left
26201..33791
/note="assembly-fragment:00014"
fragment_chain:1
33892..43982
/note="assembly-fragment:00047"
fragment_chain:1
44083..55959
/note="assembly-fragment:00350"
fragment_chain:1
56060..65631
/note="assembly-fragment:00701"
fragment_chain:1
65732..68376
/note="assembly-fragment:00007"
68477..72701
/note="assembly-fragment:00092"
72802..78093
/note="assembly-fragment:00258"
fragment_chain:2
78194..87451
/note="assembly-fragment:00602"
fragment_chain:2
clone_end:r7
vector_side:right"
BASE COUNT 26181 a 19311 c 17915 g 23233 t 811 others
ORIGIN
Query Match 15.6%: Score 62.8; DB 2; Length 87451;
Best Local Similarity 85.4%: Pred. No. 4.9e-08;
Matches 70; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 321 cagaatggggctctgtcgtatgtccagaaatggaagtgtgctatcatcagaacata 380
||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 57022 CAGACATGGGATCTCCGTGCTAGACTGGAACGACGATGCGTATTCACAGCGACGA 57081
||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 381 tcacatgcacgcagccttga 402
||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 57082 TCATCATGCACCTACGACCTTGA 57103
||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
RESULT 13
HSDJ799G3 127639 bp DNA PRI 04-APR-2001
LOCUS Human DNA sequence from clone RP4-799G3 on chromosome 1q42.11-42.3
DEFINITION Contains STS and GSSs, complete sequence.
ACCESSION AL078624
VERSION AL078624.24 GI:6272171

```

```

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 127639)
AUTHORS Bowden,P.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Nov 5, 1999 this sequence version replaced gi:6249407.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, EMBL; SWISSPROT, Tr.; TrEMBL; WPI, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
group. Further information can be found at
http://www.sanger.ac.uk/HRP/Chrl
RP4-799G3 is from the library RPCI-4 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP4-799G3 The true
right end of clone RP4-646B12 is at 100 in this sequence.
FEATURES
Source
1..127639
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.11-42.3"
/clone="RP4-799G3"
/clone_1bp="RPCI-4"
3330..3381
/note="26 copies 2 mer ag 82 conserved"
3708..4111
/note="match: GSS: Em:AQ733797"
4113..4243
/note="AluSg/X repeat: matches 2..133 of consensus"
6668..7101
/note="Tiger2a repeat: matches 1..434 of consensus"
7205..7481
/note="L1R16B repeat: matches 107..442 of consensus"
7863..8160
/note="AluSc repeat: matches 1..295 of consensus"
8868..9402
/note="match: GSS: Em:AQ588534"
9425..9518
/note="MER33 repeat: matches 19..107 of consensus"
11033..11462
/note="TIGER2 repeat: matches 2267..2718 of consensus"
11450..11515
/note="MER8 repeat: matches 174..239 of consensus"
19576..19758
/note="L1M3A repeat: matches 5832..5999 of consensus"
19985..20012
/note="14 copies 2 mer ca 100 conserved"
20013..20062
/note="25 copies 2 mer ag 82 conserved"

```


REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 168957)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 12, clone RP11-206D3
Unpublished

2 (bases 1 to 168957)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G.,
Boguslavsky, L., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lacroque, K., Lamarcas, R., Landers, T., Lebecky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Meneus, L., Milova, T., Miranda, C., Mlenka, V., Morrow, J.,
Murphy, T., Naylor, T.M., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severly, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE JOURNAL COMMENT

COMMENT

Direct Submission
Submitted (04-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02144, USA
On May 25, 2000 this sequence version replaced g1:7705174.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L7838

Center clone name: 206.D.3

Summary Statistics

Sequencing vector: M13; W7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 154625 bases at least Q40

Consensus quality: 161657 bases at least Q30

Consensus quality: 164463 bases at least Q20

Insert size: 166000; agarose-1p

Insert size: 166457; sum-of-coverage

Quality coverage: 4.0 in Q20 bases; sum-of-coverage

Quality coverage: 4.0 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
consists of 26 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
1182 1281: gap of 100 bp
1282 1920: contig of 639 bp in length
1921 2020: gap of 100 bp
2021 4480: contig of 2460 bp in length
4481 4580: gap of 100 bp
4581 6784: contig of 2204 bp in length
6785 6884: gap of 100 bp
6885 9111: contig of 2227 bp in length
9112 9211: gap of 100 bp

FEATURES

Source

9212 11337: contig of 2126 bp in length
11338 11437: gap of 100 bp
11438 14686: contig of 3249 bp in length
14687 14786: gap of 100 bp
14787 17181: contig of 2395 bp in length
17182 17281: gap of 100 bp
17282 20045: contig of 2764 bp in length
20046 20145: gap of 100 bp
20146 24869: contig of 4724 bp in length
24870 24969: gap of 100 bp
24970 30566: contig of 5597 bp in length
30567 30666: gap of 100 bp
30667 35562: contig of 4896 bp in length
35563 35662: gap of 100 bp
35663 43677: contig of 8015 bp in length
43678 43777: gap of 100 bp
43778 48684: contig of 4907 bp in length
48685 48784: gap of 100 bp
48785 55249: contig of 6465 bp in length
55250 55349: gap of 100 bp
55350 62047: contig of 6698 bp in length
62048 62147: gap of 100 bp
62148 70929: contig of 8782 bp in length
70930 71029: gap of 100 bp
71030 77706: contig of 6677 bp in length
77707 77806: gap of 100 bp
77807 84571: contig of 6765 bp in length
84572 84671: gap of 100 bp
84672 93794: contig of 9123 bp in length
93795 93894: gap of 100 bp
93895 102806: contig of 8912 bp in length
102807 102906: gap of 100 bp
102907 111867: contig of 8961 bp in length
111868 111967: gap of 100 bp
111969 123713: contig of 11746 bp in length
123714 123813: gap of 100 bp
123814 134343: contig of 10530 bp in length
134344 134443: gap of 100 bp
134444 150183: contig of 15740 bp in length
150184 150283: gap of 100 bp
150284 168957: contig of 18674 bp in length.

Location/Qualifiers

1. 168957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/map="12"
/clone="RP11-206D3"
/clone_1b="RP11 Human Male BAC"
1. 1181
/note="assembly-fragment"
1282. 1920
/note="assembly-fragment"
/note="assembly-fragment"
clone_end:17
vector_side:right"
2021. 4480
/note="assembly-fragment"
4581. 6784
/note="assembly-fragment"
6885. 9111
/note="assembly-fragment"
9212. 11337
/note="assembly-fragment"
11438. 14686
/note="assembly-fragment"
14787. 17181
/note="assembly-fragment"
17282. 20045
/note="assembly-fragment"
20146. 24869
/note="assembly-fragment"
24970. 30566
/note="assembly-fragment"


```

/note="AluB repeat: matches 199. .290 of consensus"
6913. .7099
/note="MIRFP repeat: matches 303. .477 of consensus"
repeat_region
7100. .7381
/note="AluY repeat: matches 1. .294 of consensus"
repeat_region
7382. .7439
/note="MIRFP repeat: matches 477. .536 of consensus"
misc_feature
/note="complement(7451..7674)"
repeat_region
7451. .7685
/note="match: GSS: Em:AQ721793"
/note="LTR8 repeat: matches 446. .691 of consensus"
/misc_feature
7493. .7693
/note="match: GSS: Em:AQ309297"
/misc_feature
7498. .7688
/note="match: GSS: Em:AQ349210"
/misc_feature
7509. .7678
/note="match: GSS: Em:AQ618984 Em:AQ118987 Em:AQ704859"
/misc_feature
7625. .7810
/complement(7625..7810)
/misc_feature
7686. .7837
/note="match: STR: Em:G35679"
/misc_feature
7686. .7837
/note="match: GSS: Em:AQ665254"
/misc_feature
7686. .7835
/complement(7686..7835)
/misc_feature
7687. .7859
/note="match: GSS: Em:AQ205158"
/misc_feature
7687. .7859
/note="match: GSS: Em:AQ582263"
/misc_feature
7700. .7837
/complement(7700..7837)
/misc_feature
7708. .7821
/note="match: GSS: Em:AQ352039"
/misc_feature
7741. .7993
/note="match: GSS: Em:AQ836127"
/misc_feature
7741. .7993
/complement(7741..7993)
/misc_feature
7885. .8076
/note="match: GSS: Em:AQ466383"
/misc_feature
7885. .8076
/note="LTR8 repeat: matches 1. .194 of consensus"
repeat_region
8179. .8481
/note="AluB repeat: matches 1. .307 of consensus"
repeat_region
8487. .10618
/note="AluB repeat: matches 13. .2143 of consensus"
repeat_region
10619. .10909
/note="LIP repeat: matches 13. .2143 of consensus"
repeat_region
10910. .13979
/note="AluY repeat: matches 1. .291 of consensus"
repeat_region
13980. .14214
/note="LIP repeat: matches 2143. .5220 of consensus"
repeat_region
14330. .14354
/note="AluSg repeat: matches 3. .237 of consensus"
repeat_region
14467. .14670
/note="LIP repeat: matches 4963. .4997 of consensus"
repeat_region
14845. .15243
/note="MER30 repeat: matches 1. .230 of consensus"
repeat_region
14845. .15243
/note="MLTII repeat: matches 1. .493 of consensus"
misc_feature
/complement(15351..15966)
repeat_region
16403. .16476
/note="match: GSS: Em:AQ746910"
repeat_region
16550. .16695
/note="MIR repeat: matches 103. .175 of consensus"
repeat_region
16550. .16695
/note="MIR repeat: matches 195. .240 of consensus"
repeat_region
16696. .16994
/note="AluX repeat: matches 1. .305 of consensus"
repeat_region
16995. .17142
/note="AluX repeat: matches 1. .305 of consensus"
repeat_region
16995. .17142
/note="MIR repeat: matches 47. .195 of consensus"
misc_feature
/complement(17617..18032)
repeat_region
17619. .17658
/note="match: GSS: Em:AQ813272"
repeat_region
17619. .17658
/note="match: GSS: Em:AQ813272"
repeat_region
17779. .18601
/note="20 copies 2 mer ca 85% conserved"
repeat_region
17779. .18601
/note="12 repeat: matches 1720. .2710 of consensus"
repeat_region
18676. .18832
/note="MIR repeat: matches 105. .262 of consensus"
repeat_region
18818. .18865
/note="MIR repeat: matches 160. .207 of consensus"
repeat_region
18851. .18983
/note="MIR repeat: matches 160. .207 of consensus"
repeat_region
19064. .19190
/note="MIR repeat: matches 2. .139 of consensus"
repeat_region
/note="AluX repeat: matches 2. .138 of consensus"

```

```

repeat_region      19191. 19436
/note="123 copies 2 mer ta 76% conserved"
repeat_region      19439. 19609
/note="AluX repeat: matches 135. 305 of consensus"
repeat_region      19921. 19956
/note="L2 repeat: matches 2587. 2620 of consensus"
repeat_region      19957. 20643
/note="MER67C repeat: matches 1. 710 of consensus"
repeat_region      20644. 20762
/note="L2 repeat: matches 2620. 2750 of consensus"
repeat_region      21182. 21263
/note="41 copies 2 mer tt 78% conserved"
repeat_region      21264. 21601
/note="AluX repeat: matches 3. 312 of consensus"
repeat_region      21794. 22002
/note="MR repeat: matches 16. 226 of consensus"
repeat_region      22104. 22335
/note="MR repeat: matches 14. 251 of consensus"
repeat_region      22605. 22646
/note="L2 repeat: matches 2701. 2742 of consensus"
repeat_region      22903. 23100
/note="MR repeat: matches 14. 219 of consensus"
repeat_region      23147. 23188
/note="L2 repeat: matches 2709. 2750 of consensus"
repeat_region      23480. 23761
/note="AluX repeat: matches 1. 280 of consensus"
repeat_region      23771. 24252
/note="MR1F repeat: matches 3. 502 of consensus"
repeat_region      25762. 25829
/note="LIM4 repeat: matches 5161. 5222 of consensus"
repeat_region      25830. 26119
/note="AluX repeat: matches 1. 290 of consensus"

```

Query Match	15.4%	Score 61.8	DB 9	Length 135693
Best Local Similarity	85.2%	Pred No. 1e-07		
Matches 69	Conservative 0	Mismatches 12	Indels 0	Gaps 0
QY 322	agaaatygggtctgtcgtatgtgcgccagaatggaagtcagtcgtattatcatgaagcatgat	381		
DB 99178	AGAACACGGGGTCTTGTCATATTTGCCACGAGCTGGAATGACAGTGGCTATTATCACAGGCATGAT	99119		
QY 382	catcatgcactgcagccttga	402		
DB 99118	CATGTCACATCAGACCTTAA	99098		

Search completed: January 31, 2002, 19:48:25
Job time: 1923 sec

```
Oy      322   agaaatggggtcttgcatatgtgccagaattggaagtagtgctattcataaggcatgat 381
          ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      99178 AGAGACAGGGCTTGCTATGTTGCCAGGCTGGAATGCAGTGGCTATTCAACAGGCATGAT 99119

Oy      382   catcattcaactgcagacctga 402
          ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      99118 CATGTTGCACTACAGCCTTA 99098
```

Search completed: January 31, 2002, 19:48:25
Job time: 1923 sec

us-09-641-831-5.rge

Fri Feb 1 08:56:33 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:50:07 ; Search time 67.87 Seconds
(without alignments)
1341.449 Million cell updates/sec

Title: US-09-641-831-5

Perfect score: 402

Sequence: 1 atgatgagagaccactgaaga.....atcatgactgcagccttga 402

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
4: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
5: /cgn2_6/prodata/2/lna/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/2/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48.4	12.0	87350	3	US-08-781-891-79
2	47.4	11.7	87353	3	US-08-611-587-1
3	46.6	11.6	2167	3	US-08-884-324-7
4	46.6	11.6	28994	3	US-08-884-324-14
5	44.4	11.0	3952	2	US-08-381-691-16
6	41	10.2	35060	3	US-08-814-095-7
7	39.2	9.8	7218	1	US-08-232-463-14
8	36.4	9.1	1334	3	US-08-884-324-8
9	36.4	9.1	11464	3	US-08-884-324-13
10	36.2	9.0	1696	4	US-08-124-238A-2
11	35.6	8.9	356	4	US-09-328-111-339
12	34.6	8.6	717	4	US-09-124-238A-7
13	34.6	8.6	1620	4	US-09-124-238A-32
14	34.6	8.6	1644	4	US-09-124-238A-9
15	34.6	8.6	1665	4	US-09-124-238A-33
16	34.6	8.6	1689	4	US-09-124-238A-22
17	34.4	8.6	5095	1	US-08-092-817-3
18	33.4	8.3	1597	1	US-08-166-316-1
19	33.2	8.3	1336	2	US-08-739-485-6
20	33	8.2	765	4	US-09-124-238A-21
21	32.4	8.1	1960	2	US-08-533-306A-1
22	32.4	8.1	1960	2	US-08-742-923A-1
23	32.4	8.1	2680	2	US-08-533-306A-5
24	32.4	8.1	2680	2	US-08-742-923A-5
25	32.4	8.1	2887	2	US-08-533-306A-3
26	32.4	8.1	2887	2	US-08-742-923A-3
27	32.2	8.0	4403765	4	US-09-103-840A-2

c	28	32.2	8.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl
	29	32	8.0	2162	3	US-08-948-705-5	Sequence 5, Appl
	30	32	8.0	2236	3	US-08-948-705-4	Sequence 4, Appl
	31	32	8.0	2289	3	US-08-948-705-8	Sequence 8, Appl
	32	31.8	7.9	1782	1	US-08-374-155A-13	Sequence 13, Appl
	33	31.8	7.9	1782	1	US-08-785-396-13	Sequence 13, Appl
	34	31.4	7.8	22481	5	PCT-US95-07201-43	Sequence 43, Appl
	35	31.2	7.8	646	1	US-08-370-648-1	Sequence 1, Appl
	36	31.2	7.8	646	2	US-08-531-662B-1	Sequence 1, Appl
	37	31.2	7.8	646	3	US-08-669-161A-1	Sequence 1, Appl
	38	31.2	7.8	646	3	US-08-602-039-1	Sequence 1, Appl
	39	31.2	7.8	648	1	US-08-250-162A-1	Sequence 1, Appl
	40	31.2	7.8	20303	1	US-08-370-975B-6	Sequence 6, Appl
	41	31.2	7.8	26764	1	US-08-370-975B-1	Sequence 1, Appl
	42	31	7.7	624	4	US-09-385-982-414	Sequence 414, App
	43	31	7.7	2741	1	US-08-832-877-59	Sequence 59, Appl
	44	31	7.7	2741	2	US-08-832-877-59	Sequence 59, Appl
	45	30.8	7.7	1796	1	US-07-816-283-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-781-891-79
Sequence 79, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620cendburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear
US-08-781-891-79

Query Match 12.0%, Score 48.4, DB 3, Length 87350;
Best Local Similarity 78.4%, Pred. No. 1.8e-05;
Matches 58, Conservatve 0, Mismatches 16, Indels 0;
Gy 329 gggctgtgtatgttgcacgaatggaagtgatgtgcattcatcatgagatgatacatatg 388

Comp. Genet.

DB 56638 GGGTCTGCTGTGTGCCAGGTTGGAGTGCAGTGCCTATTACAGACATGATCATACCA 56697
 QY 389 cactgagccttga 402
 DB 56698 TGCATAGCCTTGA 56711

RESULT 2
 US-08-611-587-1
 ; Sequence 1, Application US/08611587
 ; Patent No. 6150091
 ; GENERAL INFORMATION:
 ; APPLICANT: PANDOLFO, MASSIMO
 ; APPLICANT: MONTERMINI, LAURA
 ; APPLICANT: MOLTO, MARIA D.
 ; APPLICANT: Koenig, Michael
 ; APPLICANT: Campuzano, Victoria
 ; APPLICANT: Cossee, Mireille
 ; TITLE OF INVENTION: Direct Diagnosis of Friedreich Ataxia
 ; NUMBER OF SEQUENCES: 33
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fulbright & Jaworski L.L.P. Patent Dept.
 ; STREET: 1301 McKinney, Suite 5100
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: U.S.
 ; ZIP: 77010

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/611,587
 FILING DATE: 03-MAR-1996
 CLASSIFICATION: 436
 ATTORNEY/AGENT INFORMATION:
 NAME: Brashears-Macatee, Sarah J.
 REGISTRATION NUMBER: 38,087
 REFERENCE/DOCKET NUMBER: D-5901
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 713-651-5620
 TELEFAX: 713-651-5246
 TELEX: 76-2829
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8353 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapien
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT: 9q13
 UNITS: bp
 US-08-611-587-1

Query Match 11.7%; Score 47; DB 3; Length 8353;
 Best Local Similarity 78.9%; Pred. No. 1.7e-05;
 Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 QY 332 tcttgctatttccagagaatgaagtgtagtgcctatttcattagggatgatcatcagcac 391
 DB 1543 TCTTGTATGATGGCGGGCTGGAATGCTGTGCTATTTCACAGGCACATCATAGCTCAC 1602
 QY 392 tgcagccttga 402
 DB 1603 TGCAGCCTTGA 1613

RESULT 3
 US-08-884-324-7/c
 ; Sequence 7, Application US/08884324
 ; Patent No. 6060283
 ; GENERAL INFORMATION:
 ; APPLICANT: Takatori OKURA
 ; APPLICANT: Kakuji TORIGOE
 ; APPLICANT: Masahito KURIMOTO
 ; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
 ; OF INDUCING THE PRODUCTION OF INTERFERON-
 ; NUMBER OF SEQUENCES: 35
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/884,324
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 185,305/96
 ; FILING DATE: 27-JUN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BROWDY, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: OKURA-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2167 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: human
 ; TISSUE TYPE: placenta
 ; FEATURE:
 ; NAME/KEY: exon + 3'UTR
 ; LOCATION: 1..2167
 ; IDENTIFICATION METHOD: E
 ; US-08-884-324-7

Query Match 11.6%; Score 46.6; DB 3; Length 2167;
 Best Local Similarity 75.3%; Pred. No. 1.2e-05;
 Matches 58; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
 QY 323 gaaatggggctcttgctattgtgccagagaatggaagtgtagtgctatttcattagggatgatc 382
 DB 1717 GAGACGGGGCTCTCACTTTGTTCCTAGGCTCGAGTGCAGTGCCTATTTCACAGGCAAGATC 1658
 QY 383 atcatgactgcagcct 399
 DB 1657 ACATTGCACATATAGCCT 1641
 RESULT 4
 US-08-884-324-14/c
 ; Sequence 14, Application US/08884324
 ; Patent No. 6060283

GENERAL INFORMATION:
APPLICANT: Takaoori OKURA
APPLICANT: Kakui TORIGOE
APPLICANT: Masahiro KURIMOTO
TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
NUMBER OF SEQUENCES: 35
TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/884,324
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 185,305/96
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: OKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 28994 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: human
TISSUE TYPE: placenta
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..15606
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 15607..15685
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 15686..17056
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 17057..17068
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 17069..20451
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 20452..20468
IDENTIFICATION METHOD: S
NAME/KEY: mat peptide
LOCATION: 20469..20586
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 20587..21920
IDENTIFICATION METHOD: E
NAME/KEY: mat peptide
LOCATION: 21921..22054
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 22055..26827

IDENTIFICATION METHOD: E
NAME/KEY: mat peptide
LOCATION: 26828..27046
IDENTIFICATION METHOD: S
NAME/KEY: 3'UTR
LOCATION: 27047..28994
IDENTIFICATION METHOD: E
US-08-884-324-14

Query Match 11.6%; Score 46.6; DB 3; Length 28994;
Best Local Similarity 75.3%; Pred. No. 4.2e-05;
Matches 58; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 323 gaatggggtctgtctatgtccagaatggaagtgtgtgtattatagagcatgac 382
DB 28544 GAGACGGGGTCTCAGCTTGTGCTCAGGCTCGAGTGCAGTGTCTATTCACAGGCAAGTTC 28485

QY 383 atcatgcactgcagcct 399
DB 28484 ACATTGCACCTATAGCCT 28468

RESULT 5
US-08-381-691-16
Sequence 16, Application US/08381691
Patent No. 5852224
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Alpha-lac Albumin Gene Constructs
NUMBER OF SEQUENCES: 17
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/381,691
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 3952 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEetical: NO
ANTI-SENSE: NO
US-08-381-691-16

Query Match 11.0%; Score 44.4; DB 2; Length 3952;
Best Local Similarity 73.1%; Pred. No. 8.5e-05;
Matches 57; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 322 agaatgggtctgtctatgtccagaatggaagtgtgtgtattatagagcatgac 381
DB 3606 AGAGATGGGCTCTCAGCTTGTGCTCAGGCTCGAGTGCAGTGTCTATTCACAGGCTGAG 3665

QY 382 catcatgcactgcagcct 399
DB 3666 CACAGCACCTACAGCCT 3683

RESULT 6
US-08-814-095-7/c
Sequence 7, Application US/08814095
Patent No. 6025183
GENERAL INFORMATION:
APPLICANT: Soreq, Hermona
APPLICANT: Zakut, Haim
APPLICANT: Shani, Moshe
TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
ANTI-CHOLINESTERASE SUBSTANCES

us-09-641-831-6.rai

Fri Feb -1 08:56:35 2002

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: KOHN & ASSOCIATES
STREET: 30500 NO. 6025183thwestern Highway, Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/814,095
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Montgomery, Ilene N.
REGISTRATION NUMBER: 38,972
REFERENCE/DOCKET NUMBER: 2391.00066
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 35060 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Cosmid including ACHE
PROMOTOR, ACHE gene and ARS gene"
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 7q22
FEATURE:
NAME/KEY: promoter
LOCATION: 4089..22464
OTHER INFORMATION: /function= "ACHE Promotor"
OTHER INFORMATION: /standard_name= "ACHE Promotor"
FEATURE:
NAME/KEY: exon
LOCATION: 22465..22537
OTHER INFORMATION: /function= "non-translated"
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: exon
LOCATION: 24090..25177
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "(translation start:
OTHER INFORMATION: 24110)"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: exon
LOCATION: 25524..26009
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: exon
LOCATION: 27005..27274
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 4

FEATURE:
NAME/KEY: exon
LOCATION: 27255..28007
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: terminator
LOCATION: 27385..27387
FEATURE:
NAME/KEY: exon
LOCATION: 28008..28129
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: terminator
LOCATION: 28129..28131
FEATURE:
NAME/KEY: exon
LOCATION: complement (34528..34895)
OTHER INFORMATION: /function= "arsenite resistance
OTHER INFORMATION: gene"
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: exon
LOCATION: complement (34092..34358)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: exon
LOCATION: complement (33779..33963)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: exon
LOCATION: complement (33493..33591)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 4
FEATURE:
NAME/KEY: exon
LOCATION: complement (33297..33408)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: exon
LOCATION: complement (32959..33094)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: exon
LOCATION: complement (32569..32628)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 7
FEATURE:
NAME/KEY: exon
LOCATION: complement (32386..32468)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 8
FEATURE:
NAME/KEY: exon
LOCATION: complement (31894..32080)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 9
FEATURE:
NAME/KEY: exon
LOCATION: complement (31363..31534)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 10
FEATURE:

; CURRENT APPLICATION DATA:

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:56:12 ; Search time 56.12 seconds

(without alignments)
175.548 Million cell updates/sec

Title: US-09-641-831-6

Perfect score: 695

Sequence: 1 MVRTDFHKPSATLNSNTA.....CCPEKVAIHRDHHALQP.133

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID22/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	695	100.0	133	AAU00033	Novel human protei
2	541	77.8	108	AAU00032	Novel human protei
3	541	77.8	190	AAU00031	Novel human protei
4	131.5	18.9	913	AAW12522	Protein tyrosine p
5	131.5	18.9	913	AAW25156	Human PTPH1 protei
6	128.5	18.5	150	AAW43364	Human ORFX ORF128
7	128	18.4	403	AAW07095	Colon cancer assoc
8	128	18.4	481	AAW07098	Colon cancer assoc
9	128	18.4	521	AAW07099	Colon cancer assoc
10	128	18.4	652	AAW07094	Colon cancer assoc
11	128	18.4	652	AAW07050	Renal cancer assoc

12	127.5	18.3	406	19	AAW70326	Secreted protein D
13	124.5	17.9	207	22	AAE03655	Human extracellular
14	122.5	17.6	86	22	AAE55840	PDZ encoded domain
15	122.5	17.6	86	22	AAE57631	Human LIM protein
16	122.5	17.6	86	22	AAE58041	Human LIM protein
17	121.5	17.5	127	20	AAW12510	Human 5' EST secre
18	121	17.4	181	22	AAW64378	Human PDZ protein
19	117.5	16.9	316	19	AAW58397	Homo sapiens HLIM-
20	117.5	16.9	329	19	AAW58396	Homo sapiens HLIM-
21	117.5	16.9	339	21	AAE56930	Human prostate can
22	117	16.8	37	22	AAW32879	Peptide #6916 enco
23	116.5	16.8	113	21	AAW02039	Human secreted pro
24	115.5	16.6	197	22	AAW93600	Human protein sequ
25	115.5	16.6	352	22	AAW95432	Human protein sequ
26	112.5	16.2	233	20	AAW29978	Human cell junctio
27	112.5	16.2	233	22	AAU07127	Human cell junctio
28	108.5	15.6	1112	21	AAW03833	Activin receptor b
29	108.5	15.6	1161	21	AAW03832	Rat synaptic scaff
30	108.5	15.6	1277	21	AAW03149	PDZ encoded domain
31	108	15.5	97	22	AAW55836	Tax interaction pr
32	108	15.5	97	22	AAW57627	Tax interaction pr
33	108	15.5	97	22	AAW58037	Human secreted pro
34	107.5	15.5	95	21	AAW02034	Human LMP-3 (HIMP-
35	107.5	15.5	153	21	AAW36475	Human truncated bo
36	107.5	15.5	223	20	AAW97845	Human LMP-1 amino
37	107.5	15.5	223	21	AAW36473	Human LMP-2 (HIMP-
38	107.5	15.5	423	21	AAW36474	Rat bone mineralis
39	107.5	15.5	457	20	AAW97843	Human bone mineral
40	107.5	15.5	457	20	AAW97844	Rat LIM mineralisa
41	107.5	15.5	457	21	AAW36471	Human LIM minerali
42	107.5	15.5	457	21	AAW36472	Protein containing
43	104.5	15.0	80	21	AAW53231	PDZ encoded domain
44	104.5	15.0	344	22	AAW55834	Human post-synapti
45	104.5	15.0	344	22	AAW57623	

ALIGNMENTS

RESULT 1

AAU00033

ID AAU00033 standard; Protein; 133 AA.

XX AAU00033;

XX AAU00033;

DT 11-MAY-2001 (first entry)

XX Novel human protein, NHP#3.

DE Human; novel human protein; NHP#3; gene therapy; drug screening;

KW obesity; high blood pressure.

KW Homo sapiens.

OS Homo sapiens.

XX Key Location/Qualifiers

PH Misc_feature 1

FT /note= "Absent if the true start codon is

FT nucleotides 4-6 of sequence in AAU00041, the authors

FT are uncertain which is the true start codon"

XX WO200114422-A1.

XX 01-MAR-2001.

XX 18-AUG-2000; 2000WO-US22815.

XX 24-AUG-1999; 99US-0150511.

XX (LEXI-) LEXICON GENETICS INC.

XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;

XX WPI; 2001-218430/22.

DR

Query Match	100.0%	Score 695; DB 22;	Length 133;
Best Local Similarity	100.0%	Pred. No. 1.7e-76;	.
Matches 133; Conservative	0;	Mismatches 0;	Gaps 0

QY	Db	QY	Db
1	1	61	61
MAKRTTEDEPHKKSALINLSNSTATKGVYIYLEAFLECGAWGFLIKGSLHGEPLIISKVEEG	mmrttedghkpsaclinstatkgrlyyleatlegapwgfllkglehgplllskveeg	GRADPLSEKLOAGDEVVHINEVPTLSSSRKEVSLVKSQTKTLRVVRNRNVTLCCPEMKV	gkadtlskldagdevvhinevltsssrkeevslvksqtkclrlvvrnngvllccpekv
121	121	121	120
VAIHRHNDHALQF	VAIHRHNDHALQF	VAIHRHNDHALQF	VAIHRHNDHALQF
133	133	133	133
valhrhndhnlqf	valhrhndhnlqf	valhrhndhnlqf	valhrhndhnlqf

RESULT	2
AAU00032	
ID	AAU00032 standard; Protein; 108 AA

AC AAU00032;

DT 11-MAY-2001 (first entry)
 YY

DE Novel human protein, NHP#2.

KW Human; novel human protein; NHP#2; gene therapy; drug screening;
KW obesity; high blood pressure.

05 Homo sapiens.

FH	key	Location/Qualifiers
FM	W400	1

FT /note- "Absent if the true start codon is
FT nucleotides 4-6 of sequence in AA500040, the authors
FT are uncertain which is the true start codon"

PN WO200114422-A1

PD 01-MAR-2001

PF 18-AUG-2000; 2000WO-US22815.

PR 24-AUG-1999; 99US-0150511.
YY

PA (LEXI-) LEXICON GENETICS INC.
YY

~~Werner~~ CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;

XX WPI; 2001-218430/22.
DR
DR N-PSDB; AAS00040.
XX

PT Novel human polynucleotides isolated from human mammary gland cDNA
PT library, encodes novel human proteins which are useful in diagnosis
PT drug screening, clinical trial monitoring or treating behavioural
PT disorders -
XX
PS
CY Claim 4; Page 25; 29pp; English.

Claim 4; Page 25; 29pp; English.

The sequence is a novel human protein, NHP#2, which shares structural motifs with human ARLX protein. Nucleotide constructs encoding functional NHPs are used in gene therapy approaches for the modulation of NHP expression. NHP oligonucleotides can be used as hybridisation probes for screening libraries and assessing NHP gene expression patterns. Also, labeled NHP nucleotide probes can be used to screen a human genomic library. The NHP nucleotide sequences are also useful in drug screening techniques for treating symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. Examples of such manifestations may include obesity and high blood pressure.

SQ	Sequence	108 AA
----	----------	--------

Query Match	77.8%;	Score 541;	DB 22;	Length 108;
Best Local Similarity	100.0%;	Pred. No. 6	8e-58;	
Matches 107;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY 1 MMRTEDEPHSPSTLSNSNATKGRITVYEAFLEGAPGPFKKGLEHSPILITSYVEEG 60
Db 1 mrrtedfhpssatlnmentatkgrylyleaflegapgpfflkgylehspilitskveeg 60
OY 61 GKAASTSSKIOAGDEVVHNINETLSSSRKRAVSLYKGYKRLRLVVR 107
Db 61 gkaatlsskioagdevvhnineltsssrkvaavslvkgyskrlrlvvr 107

RESULT	3
AAU00031	
ID	AAU00031 standard; protein; 190 AA
XY	

AC AAU00031;

DT 11-MAY-2001 (first entry)
 YY

DE Novel human protein, NHP#1.

KW Human; novel human protein; NHP#1; gene therapy; drug screening
KW obesity; high blood pressure.
xy

OS Homo sapiens.

ET	Misc	Key	FH	Location/Qualifiers
----	------	-----	----	---------------------

FT /note= "Absent if the true start codon is
FT nucleotides 4-6 of sequence in AAS0039, the authors
FT are uncertain which is the true start codon"

PN WO200114422-A1.

PD 01-MAR-2001.

PF 18-AUG-2000; 2000WO-US22815.

PR 24-AUG-1999; 9905-0150511.
XX

PA (LEXI-) LEXICON GENETICS INC.
XX

PI Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT, XY

DR WPI; 2001-218430/22.

DR N-PSDB; AAS00031.
 XX Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX
 PS Claim 2; Page 24-25; 29pp; English.
 XX
 CC The sequence is a novel human protein, NHP#1, which shares
 CC structural motifs with human APX1 protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide probes can be used to
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 XX
 SQ Sequence 190 AA;

Query Match 77.8%; Score 541; DB 22; Length 190;
 Best Local Similarity 100.0%; Pred. No. 1.5e-57;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MMRTEDFHPSATLNSNTATKGRYYILEAFLEGGAPWGFTLKGLEHGEPLIISKVEEG 60
 DB 1 mrrttedfhpsatlnsntatkgryiyleafleggawpwtlkgglehgeplliiskveeg 60
 QY 61 GKADTLSSKLGAGDEVHINEVTLSSSRKEAVSLVKGSKYTLRLVVR 107
 DB 61 gkadtlssklqagdevvhinevtilsssrkeavslvkgskytllrlvvr 107

RESULT 4
 AAW12522
 ID AAW12522 standard; Protein; 913 AA.
 XX
 AC AAW12522;
 XX
 XX 22-MAY-1997 (first entry)
 DE Protein tyrosine phosphatase that localises to focal adhesion.
 XX
 KW Protein tyrosine phosphatase; PTPH1; focal adhesion;
 KW protein tyrosine kinase; malignancy; cancer; gene therapy;
 KW retrovirus; vector.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 PH Region 30..357
 FT /note= "N-terminal region shows homology with the
 FT /note= "N-terminal regions of the talin family"
 FT
 FT Region 357..663
 FT /note= "central region includes sequences with
 FT features of sites of phosphorylation by
 FT casein kinase II and p34cdc2"
 FT
 FT Modified-site 372
 FT /label= Phosphorylation
 FT /note= "putative p34cdc2 phosphorylation site"
 FT 381
 FT Modified-site
 FT /label= Phosphorylation
 FT /note= "putative p34cdc2 phosphorylation site"
 FT 424..428
 FT Modified-site
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT 438..442
 FT Modified-site
 FT /label= Phosphorylation

FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT 489..492
 FT Modified-site
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT 514..517
 FT Modified-site
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT 543..547
 FT Modified-site
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT 607..610
 FT Modified-site
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT 664..931
 FT Region
 FT /note= "C-terminal region shows homology to "
 XX
 XX US5959511-A.
 XX
 XX 21-JAN-1997.
 PD
 XX 14-MAR-1990; 90US-0494036.
 XX
 XX 01-MAR-1991; 91US-0663579.
 PR
 XX 14-MAR-1990; 90US-0494036.
 PR
 XX 16-AUG-1993; 93US-0107420.
 XX
 XX (COLD-) COLD SPRING HARBOR LAB.
 PA
 XX
 XX Tonks NK;
 PI
 XX WPI; 1997-107583/10.
 DR
 DR N-PSDB; AAT58627.
 XX
 XX DNA encoding protein tyrosine phosphatase - for gene therapy of
 PT cancer
 PT
 XX Disclosure; Fig 1A-B; 12pp; English.
 PS
 XX
 CC A protein tyrosine phosphatase (PTPH1) (AAW12522) catalyses the
 CC dephosphorylation of proteins in which tyrosyl residues have been
 CC phosphorylated through the action of a protein tyrosine kinase
 CC (PTK). It localises to focal adhesions, a major site of action of
 CC oncogenic PTKs. PTPH1 is the product of a cDNA clone (AAT58627)
 CC obt'd. from HeLa cells. The PTPH1 cDNA can be incorporated into a
 CC vector (pref. retroviral) to allow expression of PTPH1 in mammalian
 CC in sufficient quantities to overcome or counteract PTK activity.
 CC Phosphorylation of tyrosine residues at abnormal levels is
 CC prevented or reversed, resulting in the prevention or reversal of
 CC malignancy of cells.
 XX
 SQ Sequence 913 AA;

Query Match 18.9%; Score 131.5; DB 18; Length 913;
 Best Local Similarity 28.5%; Pred. No. 9e-07;
 Matches 35; Conservative 23; Mismatches 46; Indels 19; Gaps 4;
 QY 6 EDFHPSATLNSNTATKGRY-----IYLEAFLEGGAPWGFTLKGLEHGEPLI 53
 DB 479 ddfrh-vtkggstedsaggyckndngdsylvliripdedgkfglnkgvdkmplv 536
 QY 54 ISKVEEGKADTLSSKLGAGDEVHINEVTLSS-SRKEAVSLVKGSKYKT----LRLVVR 108
 DB 537 vsrlnpespadtclpknegdqvlvngdrdischdqvvmfikasreshrelalvrr 596
 QY 109 NGV 111
 DB 597 rav 599

RESULT 5
 ID AAY25156 standard; Protein: 913 AA.
 AAY25156.
 07-SEP-1999 (first entry)
 Human PTPH1 protein.
 PTPH1: human; protein tyrosine phosphatase; focal adhesion; cancer; localisation; treatment; overexpression; oncogenic; cell transformation; prevention; phosphotyrosine; disease; malignant.
 Homo sapiens.
 Location/Qualifiers
 30..357
 FT Region
 FT /note= "region of homology to the N-terminal domain of band 4.1, ezrin and talin. This region is known to be important for localisation to focal adhesions"
 FT Modified-site
 FT 372 /note= "potential p34cdc2 phosphorylation site"
 FT 381 /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site
 FT 424..428 /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site
 FT 434 /note= "potential casein kinase II phosphorylation site"
 FT Modified-site
 FT 438..442 /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site
 FT 489..492 /note= "potential casein kinase II phosphorylation site"
 FT Modified-site
 FT 514..518 /note= "potential casein kinase II phosphorylation site"
 FT Modified-site
 FT 543..547 /note= "potential casein kinase II phosphorylation site"
 FT Modified-site
 FT 607..610 /note= "potential casein kinase II phosphorylation site"
 FT /note= "potential casein kinase II phosphorylation site"
 US5663781-A.
 26-JAN-1999.
 04-DEC-1996; 96US-0759536.
 01-MAR-1991; 91US-0663579.
 14-MAR-1990; 90US-0494036.
 16-AUG-1993; 93US-0107420.
 04-DEC-1996; 96US-0759536.
 (COLD-) COLD SPRING HARBOR LAB.
 Tonks NK;
 WPI; 1999-131308/11.
 N-PSDB; AAY78463.
 Protein tyrosine phosphatase PTPH1 - encoded by DNA of Hela cells
 Claim 1; Fig 1A-B; 12pp; English.
 This sequence represents a novel protein tyrosine phosphatase, PTPH1, isolated from Hela cells. The protein of the invention appears to localise to focal adhesions and is therefore potentially useful in the treatment of cancer. Overexpression of PTPH1 can be used to counter the effects of oncogenic protein tyrosine kinases such as those of transforming viruses and for interfering with or reversing cell transformation. This would provide a means of preventing or reversing abnormally high levels of phosphotyrosine associated with any disease or

CC condition such as preventing or reversing malignancy associated with the
 CC activity of a protein tyrosine kinase.
 XX
 SQ Sequence 913 AA;
 Query Match 18.9%; Score 131.5; DB 20; Length 913;
 Best Local Similarity 28.5%; Pred. No. 9e-07;
 Matches 35; Conservative 23; Mismatches 46; Indels 19; Gaps 4;
 QY 6 EDFHRSATLNSMTATKGR-----IYLEAFLEGAPWFTLGGLEHGEPPI 53
 DB 479 ddfhr--vfkgsdedsqyckndngdylvllritpedgkfgfnlkygdqkmpiv 536
 QY 54 ISKVEEGKADTLSSKRLQADDEVVHNEVTLSS-SREAVSLVKGSKT-----LRLYRR 108
 DB 537 vrrlnpespadctclphnegdqlvllngldshndhqvvmfkasreshrelalvrr 596
 QY 109 NGV 111
 DB 597 rav 599
 RESULT 6
 ID AAB43364 standard; Protein: 150 AA.
 AAB43364.
 08-FEB-2001 (first entry)
 Human ORFX ORF3128 polypeptide sequence SEQ ID NO:6256.
 Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnery; antiproliferative; antiparalysant; neurotropic; neuroprotective; anticonvulsant; osteopathic; antitumor; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antineoplastic; antitumor; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.
 Homo sapiens.
 WO200058473-A2.
 05-OCT-2000.
 31-MAR-2000; 2000WO-US08621.
 31-MAR-1999; 99US-0127607.
 02-APR-1999; 99US-0127636.
 05-APR-1999; 99US-0127728.
 30-MAR-2000; 2000US-0540763.
 (CURA-) CURAGEN CORP.
 Shinkets RA, Leach M;
 WPI; 2000-602362/57.
 N-PSDB; AAC77573.
 Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -
 Claim 11; Page 5443; 5507pp; English.

XX	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnery; antispasmodic; antiparkinsonian; nootropic; immunoprotective; osteopathic; anticoagulant; antiarthritic; neuroprotective; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.	4
XX	Sequence 150 AA;	
XX	Query Match 18.5%; Score 128.5; DB 21; Length 150;	
XX	Best Local Similarity 35.9%; Pred. NO. 1.5e-07;	
XX	Matches 37; Conservative 21; Mismatches 36; Indels 9; Gaps	
QY	10 KPSATLNSWTATKG----RVLYLEAFLEGGAPWGFTLKGGLHGEPLIISKVEEGKADT 65	
DB	17 kkatvaafataseghnprvvelpktddeg--lgfninggkeqnsplrsrvipggvadr 73	
QY	66 LSSKIQAGDEVVHNEVTLSSRKE-AVSLVKGSYKTLRLVLR 107	
DB	74 -hgglkrgdqilsvngsvsvegeqhekavellkaagsgvklvvr 115	
XX	RESULT 7	
XX	AAAY07095	
XX	AAAY07095 standard; Protein; 403 AA.	
XX	AAAY07095;	
XX	02-JUL-1999 (first entry)	
DT	Colon cancer associated antigen precursor sequence.	
XX	Cancer associated antigen; diagnosis; research; treatment; human;	
XX	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
XX	prostate cancer.	
XX	Homo sapiens.	
OS	WO9904265-A2.	
PN	28-JAN-1999.	
XX	15-JUL-1998; 98WO-US14679.	
XX	22-JUN-1998; 98US-0102322.	
XX	17-OCT-1997; 97US-0896164.	
XX	10-JUL-1997; 97US-0061599.	
XX	10-OCT-1997; 97US-0061765.	
XX	10-OCT-1997; 97US-0948705.	
XX	11-OCT-1997; 97GB-0021697.	
XX	(LUDW-) LUDWIG INST CANCER RES.	
XX	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;	
XX	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;	
XX	Tureci O;	

PI Tureci O;
 XX
 DR WPI: 1999-132448/11.
 XX
 PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX
 PS Disclosure: Page 664-665; 787pp: English.
 XX
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 CC
 XX
 SQ Sequence 481 AA;
 Query Match 18.4%; Score 128; DB 20; Length 481;
 Best Local Similarity 40.0%; Pred. No. 9.4e-07;
 Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;
 QY 39 GFTKGLGEHPEPLIISVEGGRADTLSSKTLAQADVEVHINEYTLSS-SRKEAVSLVKG 97
 Db 99 glsvrglgefegcglfihlilkgqadsvg--lqygdeltvrlngyslsctheevlnlir- 155
 QY 98 SKTTLRLVRRNGVL 112
 Db 156 tkktvslkvrhlgll 170
 RESULT 9
 AAY07099 standard; Protein: 521 AA.
 ID AAY07099
 AC AAY07099;
 XX
 DT 02-JUL-1999 (first entry)
 DE Colon cancer associated antigen precursor sequence.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO9904265-A2.
 XX
 PD 28-JAN-1999.
 XX
 PF 15-JUL-1998; 98WO-US14679.
 XX
 PR 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX
 PA (LUDM-) LUDMIG INST CANCER RES.
 XX
 PI Chen Y, Gout I, Gure A, O'Hare M, Odata Y, Old LJ;

PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX
 DR WPI: 1999-132448/11.
 XX
 PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX
 PS Disclosure: Page 666-667; 787pp: English.
 XX
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 CC
 XX
 SQ Sequence 521 AA;
 Query Match 18.4%; Score 128; DB 20; Length 521;
 Best Local Similarity 40.0%; Pred. No. 1.1e-06;
 Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;
 QY 39 GFTKGLGEHPEPLIISVEGGRADTLSSKTLAQADVEVHINEYTLSS-SRKEAVSLVKG 97
 Db 99 glsvrglgefegcglfihlilkgqadsvg--lqygdeltvrlngyslsctheevlnlir- 155
 QY 98 SKTTLRLVRRNGVL 112
 Db 156 tkktvslkvrhlgll 170
 RESULT 10
 AAY07094 standard; Protein: 652 AA.
 ID AAY07094
 AC AAY07094;
 XX
 DT 02-JUL-1999 (first entry)
 DE Colon cancer associated antigen precursor sequence.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO9904265-A2.
 XX
 PD 28-JAN-1999.
 XX
 PF 15-JUL-1998; 98WO-US14679.
 XX
 PR 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX
 PA (LUDM-) LUDMIG INST CANCER RES.
 XX
 PI

Chen Y., Gout I., Gure A., O'Hare M., Obata Y., Old LJ,
Pfreundschuh M., Sahin U., Scanlan MJ., Stockert E;
Tureci O;
WPI; 1999-132448/11.
New isolated cancer associated nucleic acids and polypeptides -
isolated using sera from cancer patients, used to develop products
for the diagnosis, monitoring or treatment of cancers
Disclosure; Page 656-658; 787pp; English.

The invention relates to a method for diagnosing a disorder characterised
by expression of a human cancer associated antigen precursor coded for by
a nucleic acid molecule (NAM). The method comprises: (a) contacting a
biological sample isolated from a subject with an agent that specifically
binds to the NAM, an expression product or a fragment of an expression
product complexed with an HLA molecule; and (b) determining the
interaction between the agent and the NAM or the expression product as a
determination of the disorder. The products and methods can be used in
the diagnosis, monitoring, research, or treatment of conditions
characterised by the expression of various cancer associated antigens.
The invention provides nucleic acid sequences and encoded polypeptides
which are cancer associated antigen precursors expressed in human breast
cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
lung cancer.

Sequence 652 AA;

Query Match 18.4%; Score 128; DB 20; Length 652;
Best Local Similarity 40.0%; Pred. No. 1.5e-06;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3

QY 39 GFTLKGLGHEGPLISKVBEQGKADTLSSKLQAGDEVWHINEVTLSS-SRKEAVSLVKG 97
| : : | | | | : : : : | | : : | | : : | | : : | : : :
Db 99 glavrgglegfcglfshlkgqgdsvg--lvqgdeivringysiscstheevinlr- 155

QY 98 SYKTLRLVVRRNGVL 112
: | : : | | : : :
Db 156 tkktvsikvrhgll 170

RESULT 11
AAY07050
ID AAY07050 standard; Protein; 652 AA.
XX AC AAY07050;
XX DT
XX DT
XX DT
XX DE
XX DE
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX OS Homo sapiens.
XX OS WO9904265-A2.
XX PN
XX PD 28-JAN-1999.
XX XX 15-JUL-1998; 98WO-US14579.
XX XX 22-JUN-1998; 98US-O102322.
PR PR 17-JUL-1997; 97US-0896164.
PR PR 10-OCT-1997; 97US-0061599.
PR PR 10-OCT-1997; 97US-0061765.
PR PR 10-OCT-1997; 97US-0948705.
PR PR 11-OCT-1997; 97GB-0021697.
XX XX
XX PA (LUDW-) LUDWIG INST CANCER RES.

Fri Feb 1 08:56:35 2002

us-09-641-831-6.rag

Page 10

Query Match	Similarity	Score	DB	Length
Best Local	Similarity 35.9%	Pred. No. 3.6e-07		
Matches	28; Conservative	18; Matches	29; Indels	3; Gaps
QY	32	LEGAPWGTLLKGLLEHGEPLTISVVEEGGRKDTLSKRLACGEVYHINEVTLSS-SRKE	90	
Db	8	LYPAPWGTLLKGLLEHGEPLTISVVEEGGRKDTLSKRLACGEVYHINEVTLSS-SRKE	90	
QY	91	AVSLVKGSIKTLRLVYRR	108	
Db	66	AGNKKIGCTGSLMNLGR	83	

Search completed: January 31, 2002, 20:56:12
Job time: 80 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:57:10 ; Search time 29.66 Seconds
(without alignments)
100.908 Million cell updates/sec

Title: US-09-641-831-6
Perfect score: 695
Sequence: 1 MWRTTEFHKPSATLNSNTA.....CCPEKVVAIHRHDHALQP 133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution..

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	118.5	17.1	327	2	US-08-739-485-10
2	117.5	16.9	316	2	US-08-739-485-5
3	117.5	16.9	329	2	US-08-739-485-3
4	117	16.8	75	3	US-09-100-804-21
5	116.5	16.8	330	2	US-08-739-485-11
6	112.5	16.2	233	2	US-09-151-611-1
7	112.5	16.2	233	4	US-09-370-102-1
8	109	15.7	297	2	US-09-151-611-3
9	109	15.7	297	4	US-09-370-102-3
10	107.5	15.5	223	4	US-09-124-238A-34
11	107.5	15.5	457	4	US-09-124-238A-1
12	107.5	15.5	457	4	US-09-124-238A-10
13	104.5	15.0	454	1	US-08-166-316-2
14	103.5	14.9	73	3	US-09-100-804-28
15	97	14.0	182	3	US-09-045-632-29
16	97	14.0	283	3	US-09-045-632-27
17	97	14.0	498	3	US-09-045-632-24
18	97	14.0	502	3	US-09-045-632-20
19	97	14.0	541	3	US-09-045-632-36
20	97	14.0	599	3	US-09-045-632-28
21	97	14.0	602	3	US-09-045-632-20
22	97	14.0	642	3	US-09-045-632-35
23	97	14.0	702	3	US-09-045-632-15
24	97	14.0	818	3	US-09-045-632-25
25	97	14.0	861	3	US-09-045-632-34
26	97	14.0	918	3	US-09-045-632-21
27	97	14.0	961	3	US-09-045-632-33

28 97 14.0 1018 3 US-09-045-632-16 Sequence 16, Appl
29 97 14.0 1061 3 US-09-045-632-32 Sequence 32, Appl
30 97 14.0 1112 3 US-09-045-632-2 Sequence 2, Appl
31 97 14.0 1112 3 US-09-045-632-3 Sequence 3, Appl
32 95 13.7 505 1 US-08-123-161A-14 Sequence 14, Appl
33 95 13.7 505 1 US-08-483-278-14 Sequence 14, Appl
34 93.5 13.5 80 3 US-08-545-860D-54 Sequence 54, Appl
35 93.5 13.5 80 5 PCT-US94-04496-54 Sequence 54, Appl
36 92.5 13.3 73 3 US-09-100-804-25 Sequence 25, Appl
37 92.5 13.3 74 3 US-09-100-804-34 Sequence 34, Appl
38 90 12.9 82 3 US-09-045-632-9 Sequence 9, Appl
39 90 12.9 398 3 US-09-045-632-31 Sequence 31, Appl
40 90 12.9 441 3 US-09-045-632-37 Sequence 37, Appl
41 88.5 12.7 86 3 US-08-545-860D-53 Sequence 53, Appl
42 88.5 12.7 86 5 PCT-US94-04496-53 Sequence 53, Appl
43 87.5 12.6 75 3 US-09-100-804-22 Sequence 22, Appl
44 87.5 12.6 79 3 US-09-100-804-27 Sequence 27, Appl
45 85 12.2 2037 4 US-09-306-998-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-739-485-10
; Sequence 10, Application US/08739485
; Patent No. 5863898
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/739,485
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0142 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1020151
US-08-739-485-10

Query Match 17.1%; Score 118.5; DB 2; Length 327;

us-09-641-831-6.ra1

Fri Feb. 1 08:56:35 2002

TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN

TITLE OF INVENTION: TYROSINE PHOSPHATASES

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/100,804

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/596,291

FILING DATE: 09-AUG-1996

APPLICATION NUMBER: US 08/115,573

FILING DATE: 01-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/09943

FILING DATE: 01-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

REFERENCE/DOCKET NUMBER: LO461/7003

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 75 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-09-100-804-21

Query Match 16.8%; Score 117; DB 3; Length 75;

Best Local Similarity 38.8%; Pred. No. 2.3e-07;

Matches 26; Conservative 12; Mismatches 27; Indels 2; Gaps 2;

QY 33 EGGAPWGFTLKGLEHGEPLITSKVEGKADTLSSKLOAGDEVVHINEVTLSS-SRKEA 91

Db 2 EDGKP-GENLKGVDQKNPLVSRINSPSPADTCIPKLNEDQIVLNGRDISEHEDQV 60

QY 92 VSLVKS 98

Db 61 VMFKAS 67

RESULT 5

US-08-739-485-11

Sequence 11, Application US/08739485

Patent No. 5863898

GENERAL INFORMATION:

APPLICANT: Goli, Surya K.

APPLICANT: Hillman, Jennifer L.

APPLICANT: Bandman, Olga

TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/739,485

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0142 US

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 330 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

ORIGINAL SOURCE:

HAPLOTYPE: GenBank

IMMEDIATE SOURCE:

LIBRARY: 887580

US-08-739-485-11

Query Match 16.8%; Score 116.5; DB 2; Length 330;

Best Local Similarity 41.0%; Pred. No. 2.4e-06;

Matches 32; Conservative 9; Mismatches 34; Indels 3; Gaps 2;

QY 32 LEGAPWGFTLKGLEHGEPLITSKVEGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90

Db 7 LRGPSPWGFVLVGRDFSPAPLITISRVHAGSKA--ALAAALCPGDSIQAINGESTELMTHLE 64

QY 91 AVSLVRGSKYKTLRLVVR 108

Db 65 AQNRKRGCHDHLTSLVSR 82

RESULT 6

US-09-151-611-1

Sequence 1, Application US/09151611

Patent No. 5958731

GENERAL INFORMATION:

APPLICANT: Yue, Henry

APPLICANT: Au-Young, Janice

APPLICANT: Patterson, Chandra

TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN

FILE REFERENCE: PF-0599 US

CURRENT APPLICATION NUMBER: US/09/151,611

CURRENT FILING DATE: 1998-09-11

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PERL Program

SEQ ID NO 1

LENGTH: 233

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: 1974337

US-09-151-611-1

PRIOR AN
PRIOR F
NUMBER C
SOFTWARE

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:59:17 ; Search time 40.57 Seconds
(without alignments)
249.722 Million cell updates/sec

Title: US-09-641-831-6
Perfect score: 695
Sequence: 1 MMRTEDFHKPSATLNSNTA.....CCPEKVVAIHRDHDHALQP 133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR 68: *
2: PIR1: *
3: PIR2: *
4: PIR3: *
5: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	280	40.3	1616	I37183	gene APXL protein - human
2	131.5	18.9	913	A41109	protein-tyrosine-p
3	127.5	18.3	1171	T42372	probable guanylate
4	126	18.1	852	T10811	channel associated
5	126	18.1	870	G01974	channel associated
6	123.5	17.8	1256	JE0209	brain-specific ang
7	118.5	17.1	327	JC4385	L1M protein - rat
8	116.5	16.8	330	S71828	gene r1l protein -
9	112.5	16.2	911	I56552	synapse-associated
10	112.5	16.2	926	I44105	protein-tyrosine-p
11	111	16.0	904	I38757	homolog of Drosoph
12	111	16.0	926	T32756	homolog of Drosoph
13	110.5	15.9	317	T27179	hypothetical prote
14	108.5	15.6	1277	T14152	synaptic scaffold
15	108	15.5	1012	T23160	hypothetical prote
16	107.5	15.5	455	A55050	enigma - human
17	105.5	15.2	517	T47134	hypothetical prote
18	104.5	15.0	720	A45436	synapse-associated
19	104.5	15.0	724	JH0800	postsynaptic densi
20	104.5	15.0	767	T09599	postsynaptic densi
21	103	14.8	624	T19630	hypothetical prote
22	103	14.8	1026	T19631	hypothetical prote
23	103	14.8	1112	T32733	AMPA glutamate rec
24	101.5	14.6	431	T16191	hypothetical prote
25	99	14.2	1367	T13703	tama protein - fru
26	98.5	14.2	423	T21570	hypothetical prote
27	98.5	14.2	440	T21568	hypothetical prote
28	98.5	14.2	960	A39651	discs-large tumor
29	98.5	14.2	1387	JC5502	G-protein signalin

30	97	14.0	505	2	S62894	alpha-syntrophin -
31	96	13.8	723	2	T14765	hypothetical prote
32	95	13.7	503	2	T84771	syntrophin-1 - mou
33	95	13.7	505	2	A53214	dystrophin-associ
34	94.5	13.6	390	2	T28036	hypothetical prote
35	94.5	13.6	1131	2	T15617	hypothetical prote
36	93	13.4	538	2	I59291	betal-syntrophin -
37	90.5	13.0	628	2	T09458	numb-binding prote
38	90.5	13.0	728	2	T09457	numb-binding prote
39	88.5	12.7	1495	2	T31434	protein-tyrosine-p
40	85	12.2	2450	2	S71625	syntrophin - Pacif
41	84.5	12.2	488	2	I51379	protein tyrosine p
42	84	12.1	2294	2	I67630	protein tyrosine p
43	84	12.1	2466	2	I67629	protein-tyrosine-p
44	84	12.1	2490	1	A54971	PSD-95-related pro
45	82.5	11.9	87	2	S60315	

ALIGNMENTS

RESULT 1
I37183
gene APXL protein - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I37183
R:Schiaffino, M.V.; Bassi, M.T.; Rugarli, E.I.; Renieri, A.; Galli, L.; Ballabio, A.
Hum. Mol. Genet. 4, 373-382, 1995
A:Title: Cloning of a human homologue of the Xenopus laevis APX gene from the ocular
A:Reference number: I37183; MUID:95315933
A:Accession: I37183
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1616 <RES>
A:Cross-references: EMBL:X83543; NID:g790999; PIDN:CAA58534.1; PID:g1181628
C:Genetics:
A:Gene: GDB:APXL
A:Cross-references: GDB:582527
A:Map position: Xp22.3-Xp22.3
C:Superfamily: GLGF domain homology
F:32-104/Domain: GLGF domain homology <GLG>

Query Match 40.3%; Score 280; DB 2; Length 1616;
Best Local Similarity 52.5%; Pred. No. 1.le-19;
Matches 62; Conservative 15; Mismatches 29; Indels 12; Gaps 4;
QY 13 ATLSNNTATKGRVYILEAFLEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLQA 72
Db 14 AEATRAADGRLV--EVQLSGGAPWGFTLKGREHGEPLVITKIEGSKAAAV-DKLLA 70
QY 73 GDEVVHINEVTLSSSRKEAVSLKSGYKTLRLVVYRRNGVLLCCPEKVVAIHRDHDHA 130
Db 71 GDEIVGINDIGLSGFRQEAICLVKGSHTKLVLVRRSEL----GWR-----PHSWHA 119

RESULT 2
A41109
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 3 - human
N:Alternate names: PTPH1
C:Species: Homo sapiens (man)
C:Date: 27-Mar-1992 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C:Accession: A41109; I55698
R:Yang, Q.; Tonks, N.K.
Proc. Natl. Acad. Sci. U.S.A. 88, 5949-5953, 1991
A:Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with
A:Reference number: A41109; MUID:91296738
A:Accession: A41109
A:Molecule type: mRNA
A:Residues: 1-913 <YAN>
A:Cross-references: GB:M64572; NID:g179912; PIDN:AAA35647.1; PID:g179913
R:Ikuta, S.; Itoh, F.; Hinoda, Y.; Toyota, M.; Makiguchi, Y.; Imai, K.; Yachi, A.

J. Gastroenterol. 29, 727-732, 1994
 A:Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PRPH1 mRNA
 A:Reference number: 155698; MUID:95179278
 A:Accession: 155698
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 899-913 <RES>
 A:Cross-references: GB:S76309; NID:9913165; PIDN:AA33563.1; PID:9913166
 C:Genetics:
 A:Gene: GDB:PTPN3
 A:Cross-references: GDB:131386; OMIM:176877
 A:Map position: 9q31-9q31
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology; F
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase
 F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
 F:516-590/Domain: GLGF domain homology <GLG>
 F:670-890/Domain: protein-tyrosine-phosphatase homology <PTP>
 F:842/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:848/Binding site: substrate phosphate (Arg) #status predicted

Query Match 18.9%; Score 131.5; DB 1; Length 913;
 Best Local Similarity 28.5%; Pred. No. 4, 8e-05;
 Matches 35; Conservative 23; Mismatches 46; Indels 19; Gaps 4;

OY 6 EDHKEPATLNSMTATKGR-----IYLEAFLEGAPMGFTLKGLHEGPELI 53
 DB 479 DDFRR--VTGGSTEDASQYYCDKNDGSLYLIRTPDEKFGNKGVDQKMPLV 536
 OY 54 ISKVEEGKADTLSSKLAQDEVEHINEVTL--SRKEAVSLVKSYSKT---LRLVVR 108
 DB 537 VSRINPSPADTCIPKINBDQVILINGRDISHTHQVVMFIRKASRESRLALVIR 596

OY 109 NGV 111
 DB 597 RAV 599

RESULT 3
 T42372
 probable guanylate kinase (EC 2.7.4.8) 1, membrane-associated, splice form b - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
 C:Accession: T42372
 R:DoRosokskaya, I.; Guy, R. K.; James, G. L.
 J. Biol. Chem. 272, 31589-31597, 1997
 A:Title: Mg1-1: a membrane-associated guanylate kinase with a unique arrangement of pro
 A:Reference number: 222139; MUID:98058950
 A:Accession: T42372
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 11171 <DOB>
 A:Cross-references: EMBL:AF027503; NID:92702346; PID:92702347; PIDN:AA91995.1
 A:Experimental source: strain C57 Black/6 x CBA
 C:Genetics:
 A:Gene: Mag1-1
 C:Superfamily: WW repeat homology
 C:Keywords: alternative splicing; phosphotransferase
 F:300-337/Domain: WW repeat homology <WW1>
 F:347-384/Domain: WW repeat homology <WW2>

Query Match 18.3%; Score 127.5; DB 2; Length 1171;
 Best Local Similarity 27.3%; Pred. No. 0.00016;
 Matches 39; Conservative 25; Mismatches 52; Indels 27; Gaps 6;

OY 9 HKPS--ATLNSMTATKGR-----IYLEAFLEGAPMGFTLKGLHEGE 50
 DB 998 HAPSGOGQJQETTTTKPKQDOSEFFKGPQAQEDFYVELERGAKEGFSILRGRETYNM 1057
 OY 51 PLITKVEEGKADTLSSKLAQDEVEHINEVTLSSSRKKAVALSVKSYTLRLVVRN 109
 DB 1058 DLYVRLAEDGPAER-CGMRIGDEILINGETTKNMKSHRAIELIKNGRVRRLPLRG 1116

OY 110 GVLCCPEW-----KVAIHHRD 127
 DB 1117 DGSV--PEYAMIPPKIAMCRNE 1137

RESULT 4
 T10811
 channel associated protein of synapse 2 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
 C:Accession: T10811
 R:Riley, M.; Hata, Y.; Takai, Y.
 Submitted to the EMBL Data Library, April 1996
 A:Description: Cloning of new isoforms of PSD-95/SAP90 related genes.
 A:Reference number: 217166
 A:Accession: T10811
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-852 <IRI>
 A:Cross-references: EMBL:U53368; NID:91517939; PID:91517940
 C:Genetics:
 A:Note: PSD-95/SAP90-related gene 2 (chapsyn isoform2)
 C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
 F:198-276/Domain: GLGF domain homology <GLG>
 F:543-601/Domain: SH3 homology <SH3>
 F:663-840/Domain: guanylate kinase homology <GKI>

Query Match 18.1%; Score 126; DB 2; Length 852;
 Best Local Similarity 32.7%; Pred. No. 0.00016;
 Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;

OY 10 KPSATLNSMTATK--RYILEAFLEGAPMGFTLKGLHEGPELIISKVEEGKADTL 67
 DB 404 QPSVTLQRAISLEGEPRKVLH---KSTGLGNINIVGG-EDGEGIVSFILAGPAD-LS 458

OY 68 SKLAQDEVEHINEVTL-SSSRKKAVALSVKSYTLRLVVR 107
 DB 459 GEIQRGQILSVNGIDIRGASHQAAALAKGACQTVIIMQ 499

RESULT 5
 G01974
 channel associated protein of synapse - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 21-Jan-2000
 C:Accession: G01974
 R:Kim, E.; Cho, K.
 Submitted to the EMBL Data Library, July 1995
 A:Reference number: G08966
 A:Accession: G01974
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1870 <KIM>
 A:Cross-references: EMBL:U32376; NID:91463025; PID:91036790
 C:Genetics:
 A:Gene: chapsyn-110
 C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
 F:198-276/Domain: GLGF domain homology <GLG>
 F:543-601/Domain: SH3 homology <SH3>
 F:681-858/Domain: guanylate kinase homology <GKI>

Query Match 18.1%; Score 126; DB 2; Length 870;
 Best Local Similarity 32.7%; Pred. No. 0.00016;
 Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;

OY 10 KPSATLNSMTATK--RYILEAFLEGAPMGFTLKGLHEGPELIISKVEEGKADTL 67
 DB 404 QPSVTLQRAVALSEGEPRKVLH---KSTGLGNINIVGG-EDGEGIVSFILAGPAD-LS 458

OY 68 SKLAQDEVEHINEVTL-SSSRKKAVALSVKSYTLRLVVR 107

C:Genetics:
A:Gene: CESP:Y54G11A.10
A:Introns: 55/3; 108/3; 175/3; 228/2; 253/3
Query Match 15.9%; Score 110.5; DB 2; Length 317;
Best Local Similarity 31.0%; Pred. No. 0.0017; Indels 9; Gaps 4;
Matches 31; Conservative 24; Mismatches 36; Indels 9; Gaps 4;
QY 13 ATUNSNATKGG---RYIYLEAFLEGAPWGFTLKGGLEHGEPLIISKVEGKADTLSS 68
DB 184 ATVAFAAEGHAHPRIVELPKTDQG---LGFNVMGKQKQNSPIIYSRIIPGVADR-HG 239
QY 69 KLAGDEVVHINEVTLSSS-RKEAVSLVGSYKTLRLVVR 107
DB 240 GLKRGDOLIAVNGVNAEACHEKAVDLLKSAVGSVKLVIR 279
Query Match 15.5%; Score 108; DB 2; Length 1012;
Best Local Similarity 29.1%; Pred. No. 0.013; Indels 20; Gaps 5;
Matches 34; Conservative 24; Mismatches 39; Indels 20; Gaps 5;
QY 11 PSATLNSNTATKG---RYIYLEAF-----LEGAPWGFTLKGGLEHGEPLII 54
DB 644 PASRMRPSTTLTGFPATPNYIPLSQYNQKPSDLITVSLIRKPVGFGFRLGGVESKTPLSV 703
QY 55 SKVEGGKADTLSSKLOAGDEVVHINEVTL-SSSRKEAVSLVKS--YKTLRLVVR 108
DB 704 GOIVIGGAAEE-DGRLOEGDEIVEIDGHNVEGASHSEAVVLEAAQNKHKVLIYVR 759
Search completed: January 31, 2002, 20:59:17
Job time: 65 sec

C:Genetics:
A:Gene: CESP:Y54G11A.10
A:Introns: 55/3; 108/3; 175/3; 228/2; 253/3
Query Match 15.9%; Score 110.5; DB 2; Length 317;
Best Local Similarity 31.0%; Pred. No. 0.0017; Indels 9; Gaps 4;
Matches 31; Conservative 24; Mismatches 36; Indels 9; Gaps 4;
QY 13 ATUNSNATKGG---RYIYLEAFLEGAPWGFTLKGGLEHGEPLIISKVEGKADTLSS 68
DB 184 ATVAFAAEGHAHPRIVELPKTDQG---LGFNVMGKQKQNSPIIYSRIIPGVADR-HG 239
QY 69 KLAGDEVVHINEVTLSSS-RKEAVSLVGSYKTLRLVVR 107
DB 240 GLKRGDOLIAVNGVNAEACHEKAVDLLKSAVGSVKLVIR 279
RESULT 14
T14152
synaptic scaffolding protein S-SCAM - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Aug-2000
C:Accession: T14152
R:Hirao, K.; Hata, Y.; Ide, N.; Takeuchi, M.; Irie, M.; Yao, I.; Deguchi, M.; Toyoda, A.
J. Biol. Chem. 273, 21105-21110, 1998
A:Title: A novel multiple PDZ domain-containing molecule interacting with N-methyl-D-aspartate receptor 1
A:Reference number: Z17889; MUID:98361985
A:Accession: T14152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1277 <HR>
A:Cross-references: EMBL:AF034863; NID:93411052; PID:93411053; PIDN:RAC31124.1
C:Genetics:
A:Gene: S-SCAM
C:Function:
A:Description: may assemble receptors and cell adhesion proteins at synaptic junctions
C:Superfamily: WW repeat homology
F:307-339/Domain: WW repeat homology <WW1>
F:348-385/Domain: WW repeat homology <WW2>
Query Match 15.6%; Score 108.5; DB 2; Length 1277;
Best Local Similarity 29.8%; Pred. No. 0.015; Indels 3; Gaps 3;
Matches 25; Conservative 24; Mismatches 32; Indels 3; Gaps 3;
QY 27 YLEAFLEGAP-WGFTLKGGLEHGEPLIISKVEGKADTLSSKLOAGDEVVHIN-EVTL 84
DB 1139 YFTVDMKGAQGFSGIRGREYKMDLYRLAEDGPA-IRNGRMVGDQIIIEINGESTR 1197
QY 85 SSSRKEAVSLVGSYKTLRLVVR 108
DB 1198 DMTHARIELIAKSGGRRVLLLR 1221
RESULT 15
T23160
hypothetical protein K01A6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C:Accession: T23160
R:Cottage, A.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19701
A:Accession: T23160
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1012 <WIL>
A:Cross-references: EMBL:Z68750; PIDN:CAA92965.1; GSPDB:GN00022; CESP:K01A6.1
A:Experimental source: clone K01A6
C:Genetics:
A:Gene: CESP:K01A6.1
A:Map position: 4

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:06:28 ; Search time 26.99 Seconds
(without alignments)
180.675 Million cell updates/sec

Title: US-09-641-831-6
Perfect score: 695
Sequence: 1 MMRTEFHFKPSATLSNTA.....CCPEKVVAIHRDHHALQP 133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	280	40.3	1616	1	APXL_HUMAN
2	131.5	18.9	913	1	PTN3_HUMAN
3	126	18.1	852	1	DLG2_RAT
4	126	18.1	870	1	DLG2_HUMAN
5	118.5	17.1	327	1	PDLL_RAT
6	116.5	16.8	330	1	RIL_RAT
7	112.5	16.2	330	1	RIL_CHICK
8	112.5	16.2	911	1	DLG1_RAT
9	112.5	16.2	926	1	PTN4_HUMAN
10	111.5	16.0	849	1	DLG3_MOUSE
11	111.5	16.0	849	1	DLG3_RAT
12	111	16.0	904	1	DLG1_HUMAN
13	109.5	15.8	329	1	PDLL_HUMAN
14	104.5	15.0	724	1	DLG4_MOUSE
15	104.5	15.0	724	1	DLG4_RAT
16	104.5	15.0	767	1	DLG4_HUMAN
17	104	15.0	817	1	DLG3_HUMAN
18	103	14.8	1026	1	PTPL_CAEEL
19	101	14.5	326	1	PDLL_MOUSE
20	99	14.2	328	1	RIL_HUMAN
21	98.5	14.2	960	1	DLG1_DROME
22	98.5	14.2	1387	1	RGSC_RAT
23	93	13.4	933	1	ZO3_HUMAN
24	92.5	13.3	1447	1	RGSC_HUMAN
25	88.5	12.7	898	1	ZO3_CANFA
26	84	12.1	2485	1	PTND_HUMAN
27	83	11.9	643	1	RHOP_MOUSE
28	81	11.7	1167	1	ZO2_MOUSE
29	80.5	11.6	631	1	IL15_HUMAN
30	80	11.5	585	1	MPP3_HUMAN
31	79.5	11.4	905	1	ZO3_MOUSE
32	77	11.1	1190	1	ZO2_HUMAN
33	77	11.1	1736	1	ZO1_HUMAN

34	77	11.1	1745	1	ZO1_MOUSE
35	76	10.9	1174	1	ZO2_CANFA
36	75.5	10.9	206	1	OM25_RAT
37	75	10.8	3988	1	POLG_BVDVN
38	74.5	10.7	837	1	APBL_HUMAN
39	74.5	10.7	839	1	APBL_RAT
40	74	10.6	3898	1	POLG_BVDVS
41	72.5	10.4	145	1	OM25_HUMAN
42	72.5	10.4	621	1	TPA_STRLI
43	72	10.4	322	1	CYS1_HOMAM
44	72	10.4	397	1	DJAI_HUMAN
45	72	10.4	397	1	DJAI_MOUSE

ALIGNMENTS

RESULT 1					
ID	APXL_HUMAN	STANDARD	PRT	1616	AA
AC	Q13796;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	APICAL-LIKE PROTEIN (APXL PROTEIN).				
GN	APXL.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Retina;				
RX	MEDLINE=95315933; PubMed=7795590;				
RA	Schiaffino V.M., Bassi M.T., Rugarli E.I., Renieri A., Galli L.,				
RA	Ballabio A.;				
RT	"Cloning of a human homologue of the Xenopus laevis APX gene from the				
RT	ocular albinism type 1 critical region.";				
RL	Hum. Mol. Genet. 4:373-382(1995).				
RN	[2]				
RP	SEQUENCE OF 56-1616 FROM N.A.				
RA	Shen Y., Gibbs R.A.;				
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.				
CC	-!- TISSUE SPECIFICITY: ABUNDANT IN RETINA AND MELANOMA; ALSO IN				
CC	BRAIN, PLACENTA, LUNG, KIDNEY, PANCREAS.				
CC	-!- SIMILARITY: SOME, TO XENOPUS LAEVIS APICAL PROTEIN (APX).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X83543; CAA58534.1;				
DR	EMBL; AC002365; AAC32592.1;				
DR	HSSP; Q12959; 1PDR.				
DR	MIM; 300103;				
DR	InterPro; IPR000857; IGFBP.				
DR	InterPro; IPR001478; PDZ.				
DR	Pfam; PF00219; IGFBP; 1.				
DR	Pfam; PF00595; PDZ; 1.				
DR	SMART; SM00228; PDZ; 1.				
DR	PROSITE; PS50106; PDZ; 1.				
DR	SIMILAR 54 112				
FT	DOMAIN 150 153				
FT	DOMAIN 314 320				
FT	DOMAIN 343 346				
FT	DOMAIN 1065 1068				
FT	SEQUENCE 1616 AA; 176409 MW; 752406B5BC0B60A2 CRC64;				
SQ					

TO MOUSE SYNTROPHIN-1 AND RABBIT
DYSTROPHIN-ASSOCIATED PROTEIN.

Query Match 40.3%; Score 280; DB 1; Length 1616;
 Best Local Similarity 52.5%; Pred. No. 9,1e-20;
 Matches 62; Conservative 15; Mismatches 29; Indels 12; Gaps 4;

DB 13 ALUNSTARKGRIYIEAFLEGAPMGFTLKGLEGEPLISKEVEEGKADTLSSKLOA 72
 14 AAEETRAAGGRLV--EVLGSGAPMGFTLKGREHGEPLVITKEEGSKAAAV-DKLLA 70
 73 GDEYVHINEVTLSSRKREKSVLKSGYKTLVVRNGVLLCCPEKKVVAIHRDHHA 130
 71 GDIVGIVINDIGLSGFEQEAICLVKSGSHKTLKLVKRSEL---GMR-----PSHMA 119

DB 71 GDIVGIVINDIGLSGFEQEAICLVKSGSHKTLKLVKRSEL---GMR-----PSHMA 119

RESULT 2
 PTN3_HUMAN STANDARD; PRT; 913 AA.
 AC P26045;
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 3 (EC 3.1.3.48)
 DE (PROTEIN-TYROSINE PHOSPHATASE H) (PTP-H1).
 GN PTPN3 OR PTPH1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91296738; PubMed=1648725;
 RA Yang Q., Toks N.K.;
 RT Isolation of a cDNA clone encoding a human protein-tyrosine
 RT phosphatase with homology to the cytoskeletal-associated proteins
 RT band 4.1, ezrin, and talin.
 RT PROC. Natl. Acad. Sci. U.S.A. 88:5949-5953(1991).
 RL [2]
 RP SEQUENCE OF 194-896 FROM N.A.
 RC TISSUE-Colon;
 RX MEDLINE=9327504; PubMed=1626183;
 RA Airmura Y., Hinoda Y., Itoh F., Takekawa M., Tsujisaki M., Adachi M.,
 RA Imai K., Yachi A.;
 RT "cDNA cloning of new protein tyrosine phosphatases in the human
 RT colon."
 RL [3]
 RP Tumour Biol. 13:180-186(1992).
 RN [3]
 RP SEQUENCE OF 899-913 FROM N.A.
 RX MEDLINE=95179278; PubMed=7874267;
 RA Ikuta S., Itoh F., Hinoda Y., Toyota M., Makiguchi Y., Imai K.,
 RA Yachi A.;
 RT "Expression of cytoskeletal-associated protein tyrosine phosphatase
 RT PTPH1 mRNA in human hepatocellular carcinoma."
 RT J. Gastroenterol. 29:727-732(1994).
 RL [4]
 RP FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
 CC CYTOSKELETON.
 CC -1- CATALYTIC ACTIVITY: PROTEIN-TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN-TYROSINE + ORTHOPHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M64572; AAA35647.1;
 CC EMBL: S39392; AAB22439.2; DR

EMBL: S76309; AAB33583.1;
 DR PIR: A41109; A41109.
 DR HSSP: P18031; IPTT.
 DR MIM: 176877;
 DR InterPro: IPR000299; Band_4.1.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR000387; Tyr_phosphatase.
 DR InterPro: IPR000242; Tyr_prot_phphatase.
 DR Pfam: PF00373; Band_41; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00102; Y_phosphatase; 1.
 DR PRINTS: PRO0700; PRTYPHPTASE.
 DR PRINTS: PRO0935; BANDA1.
 DR SMART: SM00295; B41; 1.
 DR SMART: SM00228; PDZ; 1.
 DR SMART: SM00194; PTPe; 1.
 DR PROSITE: PS00660; BAND_41_1; 1.
 DR PROSITE: PS00661; BAND_41_2; 1.
 DR PROSITE: PS50057; BAND_41_3; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_PTP; 1.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 KW Structural protein; Cytoskeleton; Hydrolyase.
 FT DOMAIN 84 241
 FT DOMAIN 510 582
 FT DOMAIN 670 913
 FT ACT_SITE 842 842
 FT ACT_SITE 842 842
 SQ SEQUENCE 913 AA; 104029 MW; 29A539ACDE2P1515 CRC64;

Query Match 18.9%; Score 131.5; DB 1; Length 913;
 Best Local Similarity 28.5%; Pred. No. 2.8e-05;
 Matches 35; Conservative 23; Mismatches 46; Indels 19; Gaps 4;

DB 6 EDFHKPSATLNSMTATKGR-----TYLAFLEGAPMGFTLKGLEGEPLI 53
 479 DDFHR--VTRGSGTEDASOYCYCDKNDGSLYLITPDEGKREFNLKSGVDKMLV 536
 54 ISKVEEGKADTLSSKLOAGDEVHINEVTLSS-SRKEAVSLVKSQYKTLVVR 108
 537 VSRINPESPADTCTPRKLNEDQIVLNGRDISHTHDQVYMFKASRESRELALVIR 596

DB 109 NGV 111
 597 RAV 599

RESULT 3
 DLG2_RAT STANDARD; PRT; 852 AA.
 AC Q63622; Q62939; P70548;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSYN-110) (SYNAPTIC
 DE DENSITY PROTEIN PSD-93) (DISCS, LARGE HOMOLOG 2).
 GN DLG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96310881; PubMed=8755482;
 RA Kim E., Cho K.O., Rothschild A., Sheng M.,
 RA "Heteromultimerization and NMDA receptor-clustering activity of
 RA Chapsyn-110, a member of the PSD-95 family of proteins."
 RL Neuron 17:103-113(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96193770; PubMed=8625413;
 RA Brenman J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,

RA Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F.,
 RA Froehner S.C., Bredt D.S.;
 RT Interaction of nitric oxide synthase with the postsynaptic density
 RT protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
 RT Cell 84:757-767(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Irie M., Hata Y., Takai Y.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U49049; AAB53243.1; -
 DR EMBL: U50717; AAC52643.1; -
 DR EMBL: U53368; AAB48562.1; -
 DR HSSP: Q12959; 1PDR.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR SH3 domain; Repeat.
 FT DOMAIN 98 184 PDZ 1.
 FT 193 279 PDZ 2.
 FT DOMAIN 421 501 PDZ 3.
 FT 536 606 SH3.
 FT DOMAIN 662 852
 FT 181 182 GUANYLATE_KINASE.
 FT CONFLICT 228 228 I -> M (IN REF. 2).
 FT CONFLICT 326 326 R -> K (IN REF. 2).
 FT CONFLICT 339 339 D -> E (IN REF. 3).
 FT CONFLICT 450 454 MISSING (IN REF. 2).
 FT CONFLICT 464 465 GD -> RK (IN REF. 2).
 FT CONFLICT 474 474 D -> H (IN REF. 2).
 FT CONFLICT 476 476 R -> P (IN REF. 2).
 FT CONFLICT 478 478 A -> D (IN REF. 2).
 FT CONFLICT 484 486 AAA -> LP (IN REF. 2).
 FT CONFLICT 506 506 A -> S (IN REF. 2).
 FT CONFLICT 569 569 H -> N (IN REF. 2).
 FT CONFLICT 586 586 L -> Q (IN REF. 2).
 FT CONFLICT 626 641 MISSING (IN REF. 2).
 FT CONFLICT 639 639 K -> A (IN REF. 3).
 FT CONFLICT 726 726 F -> L (IN REF. 1).
 FT CONFLICT 733 733 N -> Y (IN REF. 2).
 FT CONFLICT 749 749 E -> V (IN REF. 1).
 FT CONFLICT 756 756 L -> H (IN REF. 2).
 FT CONFLICT 791 792 KR -> NG (IN REF. 2).
 FT CONFLICT 794 794 T -> M (IN REF. 2).
 SQ SEQUENCE 852 AA; 94934 MM; F8D414AB89C5B09 CRC64;

Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;
 QY 10 KPSATLNSNTATKG--RYVLEAFLEGAPMGFTLKGLEGEPLIISKVEEGKADTLS 67
 DB 404 OPSVTLQALNISLEGPRVVLH---KSGTGLGFNVYGS-EDGEGLFVSFTLLAGGPAD-LS 458
 QY 68 SKLQAGDEVHINEVTL-SSSRKEAVSLVKSYSKTLRLVLR 107
 DB 459 GELRGDDILSYNGIDLRGASHEDQAAALKGAGQVTITIAQ 499
 RESULT 4
 ID Dlg2_HUMAN STANDARD; PRT; 870 AA.
 AC Q15700;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSYN-110) (DISCS, LARGE
 DE HOMOLOG 2).
 GN Dlg2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96310881; PubMed=8755482;
 RA Kim E., Cho K.-O., Rothschild A., Sheng M.;
 RT "Heteromultimerization and NMDA receptor-clustering activity of
 RT Chapsyn-110, a member of the PSD-95 family of proteins.";
 RL Neuron 17:103-113(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32376; AAB04949.1; -
 DR HSSP: Q12959; 1PDR.
 DR MIM: 603583; -
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR SH3 domain; Repeat.
 FT DOMAIN 98 184 PDZ 1.
 FT 193 279 PDZ 2.
 FT DOMAIN 421 501 PDZ 3.
 FT 536 606 SH3.
 FT DOMAIN 680 870 GUANYLATE_KINASE.
 SQ SEQUENCE 870 AA; 97500 MM; 89C83BA0619F6F59 CRC64;

Query Match 18.1%; Score 126; DB 1; Length 852;
 Best Local Similarity 32.7%; Pred. No. 9e-05;

Query Match 18.1%; Score 126; DB 1; Length 870;

Best Local Similarity 32.78; Pred. No. 9.2e-05;
Matches 33; Conservative 34; Mismatches 34; Indels 8; Gaps 5;
QY 10 KPSATLNNTATKG--RYIYLEAFLEGAPMGFTLKGLEHGEPLIISKVEGKADTLS 67
Db 404 QFSMTLQRAVSLGEPRKVLH--KGSTGLGFLNVGG-EDGEGIFVSLAGPAD-LS 458
QY 68 SKLQAGDEWHINEVTL-SSSRKEAVSLKGSYKTLRLVVR 107
Db 459 GELQRGDQILSVNGIDLRGASHEQAAALKGAGTIVTIIAQ 499

RESULT 5
RIL_RAT ID PDLL_RAT STANDARD; PRT; 327 AA.
AC P52944;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
DE LIM DOMAIN PROTEIN 1) (ELFIN).
GN PDLLIM1 OR CLIM1 OR CLP36.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=96096533; PubMed=8522188;
RA Wang H., Harrison-Shostak D.C., Lemasters J.J., Herman B.;
RT "Cloning of a rat cDNA encoding a novel LIM domain protein with high
RT homology to rat RIL";
RL Gene 165:267-271(1995).
CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON (BY SIMILARITY).
CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN HEART, LUNG AND
CC LIVER, MODERATELY IN SPLEEN AND SKELETAL MUSCLE, AND AT EXTREMELY
CC LOW LEVELS (IF AT ALL) IN TESTIS AND BRAIN TISSUES.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U23769; AAA92046.1;
CC InterPro: IPR001781; LIM.
CC InterPro: IPR001478; PDZ.
CC Pfam: PF00412; LIM; 1.
CC Pfam: PF00595; PDZ; 1.
CC ProDom: PD000094; LIM; 1.
CC SMART: SM00132; LIM; 1.
CC SMART: SM00228; PDZ; 1.
CC PROSITE: PS00478; LIM_DOMAIN_1; LIM_DOMAIN_2; 1.
CC PROSITE: PS50023; LIM_DOMAIN_2; 1.
CC PROSITE: PS50106; PDZ; 1.
CC LIM domain; Metal-binding; Zinc.
CC FT DOMAIN 258 308 PDZ.
CC FT DOMAIN 327 AA; 35525 MW; 4C3B5F579FEDA4D1 CRC64;
CC SEQUENCE 327 AA; 35525 MW; 4C3B5F579FEDA4D1 CRC64;

Query Match 17.18; Score 118.5; DB 1; Length 327;
Best Local Similarity 39.24; Pred. NO. 0.00016;
Matches 31; Conservative 12; Mismatches 33; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGKADTLSKLOAGDEVVHIN-EVTLSSSRKE 90
Db 8 LQPGWGFVGLVGGKDFEQPLAISRYTPGSKA--AJANLCIGDLTAIDGEDTSSMTHLE 65
QY 91 AVSLVKGSKYKTLRLVVRN 109
Db 66 AONKIRGCVDMNTLTVSRS 84

RESULT 6
RIL_RAT ID RIL_RAT STANDARD; PRT; 330 AA.
AC P36202;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN) (RIT-18) (H-REV18).
GN RIL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=95124724; PubMed=7824279;
RA Kiess M., Scharn B., Aguzzi A., Hajnal A., Klemenz R.,
RA Schwarte-Waldhoff I., Schaefer R.;
RT "Expression of ril, a novel LIM domain gene, is down-regulated in
RT Hras-transformed cells and restored in phenotypic revertants";
RL Oncogene 10:61-68(1995).
CC -1- TISSUE SPECIFICITY: DETECTED IN SEVERAL TISSUES, MOST PROMINENT IN
CC BRAIN AND HEART OF ADULTS.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X76454; CAA53992.1;
CC PIR: S39246; S39246.
CC InterPro: IPR001781; LIM.
CC InterPro: IPR001478; PDZ.
CC Pfam: PF00412; LIM; 1.
CC Pfam: PF00595; PDZ; 1.
CC ProDom: PD000094; LIM; 1.
CC SMART: SM00132; LIM; 1.
CC SMART: SM00228; PDZ; 1.
CC PROSITE: PS00478; LIM_DOMAIN_1; LIM_DOMAIN_2; 1.
CC PROSITE: PS50023; LIM_DOMAIN_2; 1.
CC PROSITE: PS50106; PDZ; 1.
CC LIM domain; Metal-binding; Zinc.
CC FT DOMAIN 1 84 PDZ.
CC FT DOMAIN 255 305 LIM.
CC SEQUENCE 330 AA; 35521 MW; 84BEB1E629ADA584 CRC64;

Query Match 16.8; Score 116.5; DB 1; Length 330;
Best Local Similarity 41.04; Pred. No. 0.00026;
Matches 32; Conservative 9; Mismatches 34; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGKADTLSKLOAGDEVVHIN-EVTLSSSRKE 90
Db 7 LRGPSPWGFVGLVGGKDFEQPLAISRYTPGSKA--ALAAALCPGDSIQAINGESTELMTHLE 64
QY 91 AVSLVKGSKYKTLRLVVRN 108

Db 65 AONRIKCHDLITLSVSR 82

RESULT 7
RIL_CHICK STANDARD: PRT: 330 AA.
ID RIL_CHICK
AC 09PW72; 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN).
GN RIL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-White Leghorn;
RX MEDLINE=20379359; PubMed=10918612;
RA Fu S.-L., Maha A., Vogt P.K.;
RT Identification and characterization of genes upregulated in cells
transformed by v-Jun.
RL Oncogene 19:3537-3545(2000).
CC -1- FUNCTION: NEGATIVE REGULATOR OF CELL GROWTH.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF167295; AAD46655.1; -
CC InterPro: IPR001781; LIM.
CC InterPro: IPR001478; PDZ.
CC Pfam: PF00412; LIM; 1.
CC Pfam: PF00595; PDZ; 1.
CC ProDom: PD000094; LIM; 1.
CC SMART: SM00132; LIM; 1.
CC SMART: SM00228; PDZ; 1.
CC PROSITE: PS00478; LIM_DOMAIN_1; 1.
CC PROSITE: PS00023; LIM_DOMAIN_2; 1.
CC PROSITE: PS50106; PDZ; 1.
CC LIM domain; Metal-binding; zinc.
CC FT DOMAIN 8 84 PDZ.
CC FT DOMAIN 255 305 LIM.
CC FT SEQUENCE 330 AA; 35838 MW; 2C0E0B404301E404 CRC64;
SO

Query Match 16.2%; Score 112.5; DB 1; Length 330;
Best Local Similarity 38.5%; Pred. No. 0.00065;
Matches 30; Conservative 12; Mismatches 33; Indels 3; Gaps 2;

OY 32 LEGAPWGLKGLGEPLISKEVEGKADLSSKLAGDEVVHIN-EVTLSSRKE 90
DB 7 LRGPSPGKFLVGGKDFSTPLTISRINPGSKA-ALANLCPGDILLAINGESTAMTHLE 64
OY 91 AVSLVKGSKYKTLRLVRR 108
DB 65 AONRIKACVEQLLSVSR 82

RESULT 8
DLG1_RAT STANDARD: PRT: 911 AA.
ID DLG1_RAT
AC 062636;

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
DE LARGE HOMOLOG 1).
GN DLG1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RX MEDLINE=95198112; PubMed=7891172;
RA Mueller B.M., Kistner U., Voh R.W., Cases-Langhoff C., Becker B.,
RA Gundelfinger E.D., Garner C.C.;
RT Molecular characterization and spatial distribution of SAP97, a
RT novel presynaptic protein homologous to SAP90 and the Drosophila
RT discs-large tumor suppressor protein.
RL J. Neurosci. 15:2354-2366(1995).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN
CC CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO
CC THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANES BETWEEN ADJACENT
CC CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
CC -1- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY
CC SYNAPSES. IN OTHER BRAIN REGIONS, SAP97 IS FOUND IN AND ALONG
CC BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS,
CC BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A
CC VARIETY OF EPITHELIAL CELLS.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14950; AAA79976.1; -
CC HSSP: Q12958; IPDR.
CC InterPro: IPR000619; Guanylate_kin.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001452; SH3.
CC Pfam: PF00625; Guanylate_kin; 1.
CC Pfam: PF00595; PDZ; 3.
CC Pfam: PF00018; SH3; 1.
CC SMART: SM00072; GUKC; 1.
CC SMART: SM00228; PDZ; 3.
CC SMART: SM00326; SH3; 1.
CC PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
CC PROSITE: PS00052; GUANYLATE_KINASE_2; 1.
CC PROSITE: PS50106; PDZ; 3.
CC PROSITE: PS50002; SH3; 1.
CC SH3 domain; Repeat.
CC FT DOMAIN 224 310 PDZ 1.
CC FT DOMAIN 318 404 PDZ 2.
CC FT DOMAIN 465 545 PDZ 3.
CC FT DOMAIN 580 650 SH3.
CC FT DOMAIN 721 911 GUANYLATE KINASE.
CC FT DOMAIN 527 530 POLY-ALA.
CC FT SEQUENCE 911 AA; 100570 MW; 18CEBD31DD0CAFB8 CRC64;
SO

Query Match 16.2%; Score 112.5; DB 1; Length 911;
Best Local Similarity 27.9%; Pred. No. 0.0021;
Matches 34; Conservative 27; Mismatches 42; Indels 19; Gaps 5;

QY 3 RTEDHPKPSATLNSNTATKGRYIYL-EAFL-----EGAPWGTLLKGL 46
 Db 424 QSDNVHVSFSSVLGQTPASPARYSPIKAGVLGDDDEITREPKVVLHRSSTGLGFINVGG- 482
 QY 47 EGEPLIISKVEEGKADTLSSKLAQAGDEVVHINEVTL-SSRKEAVSLVKSQTKRLRV 105
 Db 483 EDGEGIFISILAGPAD-LSGELRKGDRILISVNSVDLRAASHEQAAAAALNAGQAVTI 541
 QY 106 VR 107
 Db 542 AQ 543

RESULT 9
 PTN4_HUMAN STANDARD; PRT; 926 AA.
 ID PTN4_HUMAN
 AC P29074;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 4 (EC 3.1.3.48)
 DE (PROTEIN-TYROSINE PHOSPHATASE MEG1) (PTPASE-MEG1) (MEG).
 GN PTPN4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=91288564; Pubmed=1648233;
 Gu M., York J.D., Warshawsky I., Majerus P.W.;
 RA "Identification, cloning, and expression of a cytosolic megakaryocyte
 RT protein-tyrosine-phosphatase with sequence homology to cytoskeletal
 RT protein 4.1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5867-5871(1991).
 CC -1- FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
 CC CYTOSKELETON.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2O) -
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M68941; AAA36530.1; -
 CC PIR; A41105; A41105.
 CC HSP; P18052; 1YFO.
 CC MIM; 176878; -
 CC InterPro; IPR000299; Band_4.1.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR000387; Tyr_phosphatase.
 CC InterPro; IPR000242; Tyr_prot_phphatase.
 CC Pfam; PF00373; Band_4.1.
 CC Pfam; PF00595; PDZ; 1.
 CC Pfam; PF0102; Y_phosphatase; 1.
 CC PRINTS; PR00700; PRTYPHPTASE.
 CC PRINTS; PR00935; BAND4.1.
 CC SMART; SM00295; B4.1.
 CC SMART; SM00228; PDZ; 1.
 CC SMART; SM00194; PTPC; 1.
 CC PROSITE; PS00660; BAND_4.1_1; 1.
 CC PROSITE; PS00661; BAND_4.1_2; 1.
 CC PROSITE; PS50057; BAND_4.1_3; 1.
 CC PROSITE; PS50106; PDZ; 1.
 CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.

DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 KW Structural protein; Cytoskeleton; Hydrolase.
 FT DOMAIN 84 241 BAND 4.1-LIKE.
 FT DOMAIN 517 589 PDZ.
 FT DOMAIN 677 926 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 852 852 BY SIMILARITY.
 FT ACT_SITE 852 852 BY SIMILARITY.
 SQ SEQUENCE 926 AA; 105911 MW; 4DAC6A87A675CFB0 CRC64;

Query Match 16.2%; Score 112.5; DB 1; Length 926;
 Best Local Similarity 30.0%; Pred. No. 0.0022;
 Matches 33; Conservative 22; Mismatches 36; Indels 19; Gaps 4;

QY 38 WGFTLKGLEHGEPLIISKVEEGKADTLSSKLAQAGDEVVHINEVTLSS-SRKEAVSLVK 96
 Db 528 FGFNVKGGYDQKMPVIVSRVAPGTPADLCVPRLNESGQVVLINGRDIAETHDQVVLFIK 587
 QY 97 GSYK----TLRLVVRNGV-----LLCOPEKVV-----AIHRDH 128
 Db 588 ASCERHSGELMLLVRPNVYDVVEKLENPFOYIPEKAPLDSVHQDDH 637

RESULT 10
 DLG3_MOUSE STANDARD; PRT; 849 AA.
 ID DLG3_MOUSE
 AC P70175;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC PROTEIN SAPI02 (SYNAPSE-ASSOCIATED PROTEIN 102) (DISCS,
 DE LARGE HOMOLOG 3).
 GN DLG3 OR DLGH3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=C57BL/6; TISSUE=Brain;
 RA Kohmura N., Makino S., Yagi T.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
 CC SUBUNIT NR2B.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D87117; BAAL3249.1; -
 CC HSP; Q12959; IPDR.
 CC MGD; MGI:1888986; Digh3.
 CC InterPro; IPR000619; Guanylate_kin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001452; SH3.
 CC Pfam; PF00625; Guanylate_kin; 1.
 CC Pfam; PF00595; PDZ; 3.
 CC Pfam; PF00018; SH3; 1.
 CC SMART; SM00072; Gukc; 1.
 CC SMART; SM00228; PDZ; 3.
 CC SMART; SM00326; SH3; 1.
 CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 CC PROSITE; PS00852; GUANYLATE_KINASE_2; 1.
 CC PROSITE; PS50106; PDZ; 3.
 CC PROSITE; PS50002; SH3; 1.

KW SH3 domain; Repeat. 235 PDZ 1.
 FT DOMAIN 149 PDZ 2.
 FT DOMAIN 244 PDZ 3.
 FT DOMAIN 404 PDZ 3.
 FT DOMAIN 519 SH3.
 FT DOMAIN 659 GUANYLATE KINASE.
 SO SEQUENCE 849 AA; 93482 MW; EF3EP2D7S13338EE CRC64.

Query Match 16.0%; Score 111.5; DB 1; Length 849;
 Best Local Similarity 33.0%; Pred. No. 0.0025;
 Matches 35; Conservative 23; Mismatches 37; Indels 11; Gaps 6;

OY 10 KPSATLNSNTATKGRITYLEAFLE-GGAPWGFTLKGLEH---GEP-LIISKVEEGKA 63
 DB 133 EPLSLVN--GSDGMFKYEIVLERGNSGLGFSIAGIDNPVPPDPGIFITKIIPGGA 189
 OY 64 DTLSSKLGAGDEVVHINEVTLSS-SRKEAVSLVKSYSKTLRLVVR 108
 DB 190 -AMDGRIGVNDVLRVNEVSEVVSRAVEALKEGVPVRLVVR 234

RESULT 11

ID DLG3_RAT STANDARD; PRT; 849 AA.
 AC 062936; P70547;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102) (PSD-95/SAP90 RELATED PROTEIN 1) (DISCS, LARGE HOMOLOG 3).
 GN DLG3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RC TISSUE=Brain;
 RX MEDLINE=96374358; PubMed=8780649;
 RA Mueller B.M., Kistner U., Kindler S., Chung W.J., Kuhlendahl S., Penner S.D., Lau U.-F., Voh R.W., Huganir R.L., Gundelfinger E.D., Garner C.C.;
 RT "SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes in vivo."
 RL Neuron 17:255-265(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RA Irie M., Hata Y., Takai Y.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR SUBUNIT NR2B.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U50147; AAA393031.1; -
 DR EMBL; U53367; AAB48561.1; -
 DR HSSP; Q12959; 1PDR.
 DR InterPro; IPR000619; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR Pfam; PF00595; PDZ; 3.

DR Pfam; PF00625; SH3; 1.
 DR SMART; SM00072; GuKc; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50106; PDZ; 3.
 DR PROSITE; PS50002; SH3; 1.
 DR SH3 domain; Repeat; Alternative splicing.
 FT DOMAIN 149 PDZ 1.
 FT DOMAIN 244 PDZ 2.
 FT DOMAIN 404 PDZ 3.
 FT DOMAIN 519 SH3.
 FT DOMAIN 659 GUANYLATE KINASE.
 FT VARSPLIC 627 640
 SO SEQUENCE 849 AA; 93539 MW; 34DA9C46C7BB96DB CRC64;

Query Match 16.0%; Score 111.5; DB 1; Length 849;
 Best Local Similarity 33.0%; Pred. No. 0.0025;
 Matches 35; Conservative 23; Mismatches 37; Indels 11; Gaps 6;

OY 10 KPSATLNSNTATKGRITYLEAFLE-GGAPWGFTLKGLEH---GEP-LIISKVEEGKA 63
 DB 133 EPLSLVN--GSDGMFKYEIVLERGNSGLGFSIAGIDNPVPPDPGIFITKIIPGGA 189
 OY 64 DTLSSKLGAGDEVVHINEVTLSS-SRKEAVSLVKSYSKTLRLVVR 108
 DB 190 -AMDGRIGVNDVLRVNEVSEVVSRAVEALKEGVPVRLVVR 234

RESULT 12

ID DLG1_HUMAN STANDARD; PRT; 904 AA.
 AC 012959; Q12958;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS, LARGE HOMOLOG 1).
 GN DLG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95024052; PubMed=7937897;
 RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;
 RT "Cloning and characterization of hdlg: the human homologue of the Drosophila discs large tumor suppressor binds to protein 4.1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
 RX MEDLINE=9638231; PubMed=8757139;
 RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F., Marfatia S.M., Chishti A.H., Liddington R.C.;
 RT "Crystal structure of a PDZ domain."
 RL Nature 382:649-652(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNTS. ASSOCIATES WITH PROTEIN 4.1.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2. ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```

CC CC EMBL; UI3897; AAA50599.1; -
CC CC EMBL; UI3896; AAA50598.1; -
CC CC PDB; 1PDR; 23-JUL-97.
CC CC MIM; 601014; -.
CC CC InterPro; IPR000619; Guanylate_kin.
CC CC InterPro; IPR001478; PDZ.
CC CC InterPro; IPR001452; SH3.
CC CC Pfam; PF00625; Guanylate_kin; 1.
CC CC Pfam; PF00595; PDZ; 3.
CC CC Pfam; PF00018; SH3; 1.
CC CC SMART; SM00072; GUKC; 1.
CC CC SMART; SM00228; PDZ; 3.
CC CC SMART; SM00326; SH3; 1.
CC CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
CC CC PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
CC CC PROSITE; PS0106; PDZ; 3.
CC CC PROSITE; PS00002; SH3; 1.
CC CC KW SH3 domain; Repeat; Alternative splicing; 3D-structure.
CC CC FT DOMAIN 224 310 PDZ 1.
CC CC FT DOMAIN 319 405 PDZ 2.
CC CC FT DOMAIN 466 546 PDZ 3.
CC CC FT DOMAIN 581 651 SH3.
CC CC FT DOMAIN 714 904 GUANYLATE KINASE.
CC CC FT VARSPLIC 669 680 EIPDDMGSKGLK -> QSFNDRKKNLFSRPFYKKNQDS
CC CC FT EQUESTADOQ (IN ISOFORM 2).
CC CC FT SEQUENCE 904 AA; 100354 MW; B78798D6B0920D4 CRC64;

Query Match 16.0%; Score 111; DB 1; Length 904;
Best Local Similarity 30.3%; Pred. No. 0.003;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

QY 7 DFH-KPSATLNSNTAKGRYIVL-BAFL-----EGGAPWGFTLKGLGCHG 49
DB 428 DNHVSFSLGQTPASPARSPVSKAVLGLDDEITREPKVVLHRSSTGLGFNVIGG-EDG 486
QY 50 EPLIISKVEGKADTLSSKLQAGDEVHINEVTL-SSSRKEAVSLVSGSYKTLRLVVR 107
DB 487 EGIFISFILAGGPAD-LSGELRKGRDRIISVNSVDLRAASHEQAAAALKNAGQAVTVAQ 544

RESULT 13
ID PDLL_HUMAN STANDARD; PRT; 329 AA.
AC 000151;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
DE LIM DOMAIN PROTEIN 1) (ELFIN).
GN PDLLIM1 OR CLIM1 OR CLP36.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=99145108; PubMed=10022510;
RA Kotaka M., Ngai S.M., Garcia-Barcelo M., Tsui S.K.W., Fung K.P.,
RA Lee C.Y., Waye M.M.Y.;
RT "Characterization of the human 36-kDa carboxyl terminal LIM domain
RT protein (hCLIM1)".
RL J. Cell. Biochem. 72:279-285(1999).
RN [2]
RP INTERACTION WITH ALPHA-ACTININ 2.
RX MEDLINE=20320865; PubMed=10861853;
RA Kotaka M., Kustin S., Ngai S., Chan K., Lau Y., Lee S.M., Li H.Y.,
RA Ng E.K., Schaper J., Tsui S.K.W., Fung K.P., Lee C.Y., Waye M.M.Y.;
RT "Interaction of hCLIM1, an enigma family protein, with alpha-actinin

```

```

RT 2.;
RL J. Cell. Biochem. 78:558-565(2000).
RN [3]
RP INTERACTION WITH ALPHA-ACTININS 1 AND 4.
RX MEDLINE=20219155; PubMed=10753915;
RA Vallienus T., Luukko K., Makela T.P.;
RT "CLP-36 PDZ-LIM protein associates with nonmuscle alpha-actinin-1 and
RT alpha-actinin-4".
RL J. Biol. Chem. 275:11100-11105(2000).
CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON.
CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININ.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN THE HEART AND SKELETAL
CC MUSCLE, MODERATELY EXPRESSED IN THE LIVER, SMALL INTESTINE,
CC COLON, PLACENTA, AND LUNG. A LOWER LEVEL EXPRESSION IS SEEN IN
CC LIVER, THYMUS, KIDNEY, PROSTATE AND PANCREAS AND IS NOT FOUND IN
CC THE BRAIN, TESTIS, OVARY, AND PERIPHERAL BLOOD LEUKOCYTES.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```

CC CC EMBL; U90878; AAC05580.1; -
CC CC MIM; 605900; -
CC CC InterPro; IPR001781; LIM.
CC CC InterPro; IPR001478; PDZ.
CC CC Pfam; PF00412; LIM; 1.
CC CC Pfam; PF00595; PDZ; 1.
CC CC ProDom; PD000094; LIM; 1.
CC CC SMART; SM00132; LIM; 1.
CC CC SMART; SM00228; PDZ; 1.
CC CC PROSITE; PS00478; LIM_DOMAIN_1; FALSE_NEG.
CC CC PROSITE; PS00023; LIM_DOMAIN_2; 1.
CC CC PROSITE; PS0106; PDZ; 1.
CC CC CytoSkeletoN; LIM domain; Metal-binding; Zinc.
CC CC DOMAIN 3 84 PDZ.
CC CC DOMAIN 260 310 LIM.
CC CC SEQUENCE 329 AA; 36171 MW; 6520F1A932CFA312 CRC64;

```

```

Query Match 15.8%; Score 109.5; DB 1; Length 329;
Best Local Similarity 38.0%; Pred. No. 0.0013;
Matches 30; Conservative 12; Mismatches 34; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLGHEGEPLIISKVEGGRADTLSSKLQAGDEVHIN-EVTLSSSRKE 90
DB 8 LQGPCPGWFRVLGKDFEQPLAISRVTPGSKA--ALANLCIGDVTIDGENTSNMTHLE 65
QY 91 AVSLVKGYSYKTLRLVVRN 109
DB 66 AQNRKIGCTDNLTLTVARS 84

```

```

RESULT 14
ID DLG4_MOUSE STANDARD; PRT; 724 AA.
AC 062108;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC DENSITY PROTEIN 95 (PSD-95) (PRESYNAPTIC PROTEIN SAP90)
DE (SYNAPSE-ASSOCIATED PROTEIN 90) (DISCS, LARGE HOMOLOG 4).
GN DLG4 OR DLGH4 OR PSD95.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

```


Search completed: January 31, 2002, 21:06:29
Job time: 52 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:05:35 ; Search time 72.88 Seconds
(without alignments)
266,935 Million cell updates/sec

Title: US-09-641-831-6
Perfect score: 695
Sequence: 1 MKRTTEDEHKPSATLNSNTA.....CCPEKKVYAIHRHDHALQP 133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL.17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	415	59.7	1986	11	Q9QXNO mus musculu
2	273	39.3	1502	4	Q9ULI8 homo sapien
3	145	20.9	1454	4	Q9H4G2 Q9H4G2 homo sapien
4	134.5	19.4	298	13	Q9P046 Q9P046 gallus galli
5	132.5	19.1	315	13	Q9P047 gallus galli
6	130	18.7	493	5	Q9W443 drosophila
7	128	18.4	39	4	Q9H4T0 homo sapien
8	128	18.4	403	4	Q9UPC3 Q9UPC3 mus musculu
9	128	18.4	548	11	Q9ES65 mus musculu
10	128	18.4	552	4	Q9UM04 Q9UM04 homo sapien
11	128	18.4	652	4	Q9Y6N9 Q9Y6N9 homo sapien
12	128	18.4	910	11	Q9ES64 mus musculu
13	127.5	18.3	400	5	Q9V7F2 Q9V7F2 drosophila
14	127.5	18.3	1171	11	Q9A893 mus musculu
15	126	18.1	521	4	Q9UM17 Q9UM17 homo sapien
16	124.5	17.9	207	4	Q9HAP6 Q9HAP6 homo sapien
17	123.5	17.8	207	11	Q9Z252 ratu mus musculu
18	123.5	17.8	207	11	Q9H951 mus musculu
19	123.5	17.8	234	11	Q9D2Q1 mus musculu

20	123.5	17.8	1256	4	Q75085	Q75085 homo sapien
21	123	17.7	171	5	Q17458	Q17458 schistosoma
22	122.5	17.6	596	4	Q60705	Q60705 homo sapien
23	121	17.4	202	4	Q9NTP3	Q9NTP3 homo sapien
24	121	17.4	517	4	Q9H5P4	Q9H5P4 homo sapien
25	120	17.3	926	11	Q9W022	Q9W022 mus musculu
26	118.5	17.1	138	11	Q9D130	Q9D130 mus musculu
27	118.5	17.1	288	11	Q9R021	Q9R021 mus musculu
28	118.5	17.1	288	11	Q9WVH1	Q9WVH1 mus musculu
29	117.5	16.9	199	4	Q9BVP4	Q9BVP4 homo sapien
30	117.5	16.9	214	11	Q9CRA2	Q9CRA2 mus musculu
31	117.5	16.9	239	11	Q9QYR1	Q9QYR1 mus musculu
32	117.5	16.9	315	13	Q98R15	Q98R15 salmo salar
33	117.5	16.9	327	11	Q98K93	Q98K93 mus musculu
34	117.5	16.9	329	4	Q9BP29	Q9BP29 homo sapien
35	117.5	16.9	337	11	Q9QYNO	Q9QYNO mus musculu
36	117.5	16.9	364	4	Q43580	Q43580 homo sapien
37	117.5	16.9	591	11	Q62820	Q62820 ratu mus musculu
38	117.5	16.9	591	11	Q9QYR2	Q9QYR2 mus musculu
39	116.5	16.8	283	4	Q9Y4Z5	Q9Y4Z5 homo sapien
40	116.5	16.8	470	4	Q9Y4Z3	Q9Y4Z3 mus musculu
41	116.5	16.8	617	4	Q9Y4Z4	Q9Y4Z4 homo sapien
42	116.5	16.8	661	11	Q9JRS3	Q9JRS3 mus musculu
43	116.5	16.8	723	11	Q9JRS4	Q9JRS4 mus musculu
44	116.5	16.8	723	11	Q9WVH2	Q9WVH2 mus musculu
45	116.5	16.8	734	4	Q75112	Q75112 homo sapien

ALIGNMENTS

RESULT 1
ID Q9QXNO PRELIMINARY: PRT: 1986 AA.
AC Q9QXNO:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PDZ DOMAIN ACTIN BINDING PROTEIN SHROOM.
GN SHRM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20055594; PubMed=10589677;
RA Hildebrand J.D., Soriano P.;
RT "Shroom, a PDZ domain-containing actin-binding protein, is required
for neutral tube morphogenesis in mice."
RL Cell 99:485-497(1999).
DR EMBL: AF199421; AAF13269.1;
DR HSPF: P29476; IOAV.
DR MGD: MGI:1351655; Shrm.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
SQ SEQUENCE 1986 AA: 215261 MW: 18056FCGB772EABF CRC64;

Query Match 59.7%; Score 415; DB 11; Length 1986;
Best local similarity 75.0%; Pred. No. 8.3e-31;
Matches 81; Conservative 12; Mismatches 15; Indels 0; Gaps 0;
QY 2 MKRTTEDEHKPSATLNSNTATKRYITLFAFLBSGAGWGTLLKGLGHEGPIITISKYEKG 61
1 MKRTTEDEHKPSATLNSNTATKRYITLFAFLBSGAGWGTLLKGLGHEGPIITISKYEKG 60
QY 62 KADTSSKLOAGEVYHINEVTLSSRKAVSLVKSYSYTLRLVVRN 109
61 KADTSSKLOAGEVYHINEVTLSSRKAVSLVKSYSYTLRLVVRN 108

Purification and characterization of an alpha-actinin-binding PDZ-LIM protein that is up-regulated during muscle differentiation.
 RT J Biol. Chem. 274:29242-29250(1999).
 CC -1- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC IONS.
 CC EMBL: AJ29218; CAB53970.1; -
 DR InterPro: IPR001781; LIM.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00412; LIM; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR ProDom: PD000094; LIM; 1.
 DR SMART: SM00328; PDZ; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
 DR PROSITE: PS50106; PDZ; 1.
 KM LIM motif: Metal-binding; Zinc.
 SO SEQUENCE 315 AA; 34365 MW; 86414A0A89BCA35 CRC64;

Query Match 19.1%; Score 132.5; DB 13; Length 315;
 Best Local Similarity 36.8%; Pred. No. 8.6e-05;
 Matches 32; Conservative 16; Mismatches 36; Indels 3; Gaps 2;

QY 32 LEGAPMGFTLKGLEHGEPLIISKEGKADTSSKLGADDEVYHINEV-TLSSSRE 90
 Db 7 LPGAPEWGFRLSGIDENQPLITRTPGSKAST--ANLCREDITVAINGLSTEWATIND 64
 QY 91 AVSLVKGSTKTLRYVRNNGVLLCP 117
 Db 65 AOEKKAHAHQLSLRERAKETKMSPO 91

RESULT 6
 Q9M443 PRELIMINARY; PRT; 493 AA.

AC Q9M443
 DT 01-MAR-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG5921 PROTEIN.
 GN CG5921
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandas P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Bell G., Nelson C.R., Miklos G.L.G.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Abdayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busan D.A., Butler H., Caddieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.T., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003436; AAF46116.1; -
 DR HSSP: Q12959; 1PDR.
 DR FlyBase: FBgn0029835; CG5921.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 2.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 SO SEQUENCE 493 AA; 50574 MW; C46711C9F19797A4 CRC64;

Query Match 18.7%; Score 130; DB 5; Length 493;
 Best Local Similarity 37.8%; Pred. No. 0.00026;
 Matches 37; Conservative 16; Mismatches 35; Indels 10; Gaps 5;

QY 33 EGAPMGFTLKGLEHGEPLIISKEGKADTSSKLGADDEVYHINEVTLSSS-RKA 91
 Db 80 ERGSTYGFVGRGREGHGFVSHVEHGGEAHL--KGIRIGDQILIRINGRLDAVHKEF 137
 QY 92 VSLVKGSTKTLRYVRNNGVLLC--CPE---WKYVAI 123
 Db 138 IOLVAGODR-VTLKVGWGLPVRDLPERRLSWSVKL 174

RESULT 7
 Q9H4T0 PRELIMINARY; PRT; 39 AA.

AC Q9H4T0
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE BA554P16.1 (KIAA1202) (FRAGMENT).
 GN BA554P16.1
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bird C.;
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL359272; CAC12948.1; -
 FT NON_TER 39
 SO SEQUENCE 39 AA; 4260 MW; 2773DAE613D25D3F CRC64;

Query Match 18.4%; Score 128; DB 4; Length 39;
 Best Local Similarity 67.6%; Pred. No. 1.6e-05;
 Matches 23; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 23 GRTYVLEAFLEGAPMGFTLKGLEHGEPLIISK 56
 Db 6 GSFQYVPVQLOGAPMGFTLKGLEHGEPLTVSK 39

RESULT 8
 Q9UPC3 PRELIMINARY; PRT; 403 AA.

ID Q9UPC3
 AC Q9UPC3;

Fri Feb 1 08:56:37 2002

01-MAY-2000 (TREMBlrel. 13, Created)
01-MAY-2000 (TREMBlrel. 13, Last sequence update)
01-JUN-2001 (TREMBlrel. 17, Last annotation update)
ANTIGEN NY-CO-37.
NY-CO-38.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP TISSUE-COLON CANCER METASTASIS TO LUNG;
RX MEDLINE=9627252; PubMed=9610721;
RA Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
Gordan J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT "Characterization of human colon cancer antigens recognized by
autologous antibodies."
RL Int. J. Cancer 76:652-658(1998).
DR EMBL; AF039699; AAC18048.1; .
DR HSSP; P29476; 1QAV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
SQ SEQUENCE 403 AA; 45838 MW; AB24FB6E1D3144FD CRC64;

Query Match 18.4%; Score 128; DB 4; Length 403;
Best Local Similarity 40.0%; Pred. No. 0.00032;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;

QY 39 GFTLKGLEHGEPLIISKVEGGKADTLSSKLQAGDEVVHINEVTLS-SRKEAVSLVK 97
| : : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
DB 99 GLSVRGLEFGCGLFISHIKGGQADSVG--LQVGDEIVRINGYSISSCTHEEVINLIR- 155
: : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
QY 98 SYKTLRLVVRNGVL 112
: : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
DB 156 TKKTVSIKVRHIGLI 170

RESULT 9
Q9ES65 PRELIMINARY; PRT; 548 AA.
AC Q9ES65;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HARMONIN ISOFORM A1.
GN 2010016F01RIK OR USHIC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20428180; PubMed=10973247;
RA Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,
Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.;
RT "A defect in harmonin, a PDZ domain-containing protein expressed in
the inner ear sensory hair cells, underlies Usher syndrome type 1C."
RL Nat. Genet. 26:51-55(2000).
DR EMBL; AF28924; AAG12457.1; .
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 548 AA; 61712 MW; 44A245AA32091070 CRC64;

Query Match 18.4%; Score 128; DB 11; Length 548;
Best Local Similarity 40.0%; Pred. No. 0.00047;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;

QY 39 GFTLKGLEHGEPLIISKVEGGKADTLSSKLQAGDEVVHINEVTLS-SRKEAVSLVK 97
| : : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
DB 99 GLSVRGLEFGCGLFISHIKGGQADSVG--LQVGDEIVRINGYSISSCTHEEVINLIR- 155
: : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
QY 98 SYKTLRLVVRNGVL 112
: : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
DB 156 TKKTVSIKVRHIGLI 170

RESULT 10
Q9UM04 PRELIMINARY; PRT; 552 AA.
AC Q9UM04;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE AUTOIMMUNE ENTEROPATHY-RELATED ANTIGEN AIE-75.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SMALL INTESTINE;
RX MEDLINE=99431904; PubMed=10500064;
RA Kobayashi I., Imamura K., Kubota M., Ishikawa S., Yamada M.,
Tonoki H., Okano M., Storch W.F., Moriuchi T., Sakiyama Y.,
Kobayashi K.;
RT "Identification of an autoimmune enteropathy-related 75-kilodalton
antigen."
RL Gastroenterology 117:823-830(1999).
DR EMBL; AB018687; BAA81740.1; .
DR HSSP; P29476; 1QAV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 552 AA; 62210 MW; 7E75CEEB73C57F41 CRC64;

Query Match 18.4%; Score 128; DB 4; Length 552;
Best Local Similarity 40.0%; Pred. No. 0.00047;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;

QY 39 GFTLKGLEHGEPLIISKVEGGKADTLSSKLQAGDEVVHINEVTLS-SRKEAVSLVK 97
| : : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
DB 99 GLSVRGLEFGCGLFISHIKGGQADSVG--LQVGDEIVRINGYSISSCTHEEVINLIR- 155
: : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
QY 98 SYKTLRLVVRNGVL 112
: : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
DB 156 TKKTVSIKVRHIGLI 170

RESULT 11
Q9Y6N9 PRELIMINARY; PRT; 652 AA.
AC Q9Y6N9;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANTIGEN NY-CO-38.
GN NY-CO-38.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98272252; PubMed=9610721;
RA Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
Gordan J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT "Characterization of human colon cancer antigens recognized by

Match	Conservative	Mismatches	Indels	Gaps
29	15	28	3	95

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: January 31, 2002, 21:06:28 ; Search time 26.99 Seconds
(without alignments)
146.714 Million cell updates/sec

Title: US-09-641-831-4

Perfect score: 545
Sequence: 1 MMRTDEPHKPSATLNSNTA.....KEAVSLVKGSKYKRLRYVRS 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270.5	49.6	1616	1 APXL_HUMAN	Q13796 homo sapien
2	126	23.1	852	2 DLG2_RAT	O63622 rattus norv
3	126	23.1	870	1 DLG3_HUMAN	Q15700 homo sapien
4	122.5	22.5	913	1 PTN3_HUMAN	P26045 homo sapien
5	113.5	20.8	327	1 PDLL_RAT	P52944 rattus norv
6	112.5	20.6	330	1 RIL_RAT	P36202 rattus norv
7	112.5	20.6	911	1 DLG1_RAT	O62696 rattus norv
8	111	20.4	904	1 DLG1_HUMAN	Q12959 homo sapien
9	108.5	19.9	330	1 RIL_CHICK	O9PW72 gallus gall
10	106.5	19.5	849	1 DLG3_MOUSE	P70175 mus musculu
11	106.5	19.5	849	1 DLG3_RAT	O62936 rattus norv
12	104.5	19.2	329	1 PDLL_HUMAN	O00151 homo sapien
13	104.5	19.2	724	1 DLG4_MOUSE	O62108 mus musculu
14	104.5	19.2	724	1 DLG4_RAT	P31016 rattus norv
15	104.5	19.2	767	1 DLG4_HUMAN	P78332 homo sapien
16	99.5	18.3	817	1 DLG3_HUMAN	O92786 homo sapien
17	99	18.2	1026	1 PTN1_CAEL	P28191 caenorhabd1
18	98.5	18.1	926	1 PTN4_HUMAN	P29074 homo sapien
19	98.5	18.1	960	1 DLG1_DROME	P31007 dirosophila
20	98.5	18.1	1387	1 RGSC_RAT	O08774 rattus norv
21	96	17.6	326	1 PDLL_MOUSE	O70400 mus musculu
22	95	17.4	328	1 RIL_HUMAN	P50479 homo sapien
23	92.5	17.0	1447	1 RGSC_HUMAN	O14924 homo sapien
24	83.5	15.3	898	1 Z03_CANFA	O62663 canis fam11
25	83.5	15.3	933	1 Z03_HUMAN	O95049 homo sapien
26	83	15.2	643	1 PHOP_MOUSE	O61085 mus musculu
27	82.5	15.1	2485	1 PTND_HUMAN	Q12923 homo sapien
28	80	14.7	585	1 MPB3_HUMAN	Q13368 homo sapien
29	79.5	14.6	905	1 Z03_MOUSE	O9QXY1 mus musculu
30	78	14.3	631	1 IL16_HUMAN	Q14005 homo sapien
31	73.5	13.5	3988	1 POLG_BVDVN	P19711 bovine vira
32	72.5	13.3	3898	1 POLG_BVDVN	O01499 bovine vira
33	72	13.2	145	1 OM25_HUMAN	P57105 homo sapien

34	72	13.2	1190	1 Z02_HUMAN	O9UDY2 homo sapien
35	72	13.2	1736	1 Z01_HUMAN	O07157 homo sapien
36	72	13.2	1745	1 Z01_MOUSE	P39447 mus musculu
37	71.5	13.1	3898	1 POLG_HCVB	P19712 hog cholera
38	71.5	13.1	3898	1 POLG_HCVB	P21530 hog cholera
39	71	13.0	322	1 CYS1_HUMAN	P13277 homarus ame
40	71	13.0	466	1 EM55_MOUSE	P70290 mus musculu
41	71	13.0	1167	1 Z02_MOUSE	O95001 mus musculu
42	71	13.0	1174	1 Z02_CANFA	O95168 canis fam11
43	70	12.8	206	1 OM25_RAT	O9VWJ4 rattus norv
44	69.5	12.8	1427	1 ABC1_SCHPO	O92337 schizosacch
45	69	12.7	1591	1 TIMM_HUMAN	Q13009 homo sapien

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	1616 AA.
APXL_HUMAN	APXL_HUMAN			
AC	Q13796;			
DT	01-NOV-1997 (rel. 35, Created)			
DT	01-NOV-1997 (rel. 35, Last sequence update)			
DT	20-AUG-2001 (rel. 40, Last annotation update)			
DE	APICAL-LIKE PROTEIN (APXL PROTEIN).			
GN	APXL			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Retina;			
RX	MEDLINE=95315933; PubMed=77955590;			
RA	Schaffino V.M., Bassi M.T., Rugaril E.I., Renieri A., Galli L.,			
RA	Ballabio A.;			
RT	"Cloning of a human homologue of the Xenopus laevis Apx gene from the			
RT	ocular albinism type 1 critical region.";			
RL	Hum. Mol. Genet. 4:373-382(1995).			
RN	[2]			
RP	SEQUENCE OF 56-1616 FROM N.A.			
RA	Shen Y., Glabbs R.A.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
CC	- TISSUE SPECIFICITY: ABUNDANT IN RETINA AND MELANOMA; ALSO IN			
CC	BRAIN, PLACENTA, LUNG, KIDNEY, PANCREAS			
CC	- SIMILARITY: SOME, TO XENOPUS LAEVIS APICAL PROTEIN (APX).			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: X83543; CAAS8534.1; -			
DR	EMBL: AC002365; AAC32592.1; -			
DR	HSSP: Q12959; IPDR.			
DR	MIM: 300103; -			
DR	InterPro: IPR000867; IGFBP.			
DR	InterPro: IPR001478; PDZ.			
DR	Pfam: PF00595; PDZ. 1.			
DR	Pfam: PF00595; PDZ. 1.			
DR	SMART: SM00228; PDZ. 1.			
DR	PROSITE: PS50106; PDZ. 1.			
FT	SMILAR	54	112	
FT				TO MOUSE SYNTHROPIN-1 AND RABBIT
FT	DOMAIN	150	153	DYSTROPHIN-ASSOCIATED PROTEIN.
FT	DOMAIN	314	320	POLY-SER.
FT	DOMAIN	343	346	POLY-PRO.
FT	DOMAIN	1065	1068	POLY-ALA.
FT	DOMAIN	1616	16409	POLY-PRO.
SQ	SEQUENCE	1616 AA; 176409 MW; 752406B5BC0B60A2 CRC64;		

Query Match 49.6%; Score 270.5; DB 1; Length 1616;
Best Local Similarity 58.9%; Pred. No. 3.3e-19;
Matches 56; Conservative 14; Mismatches 22; Indels 3; Gaps 2;

QY 13 ATLSNTATKGRVYLEAFLEGGAPWGFTLKGLEHGEPLIIISKVEEGKADTLSSKLOA 72
DB 14 AEAETRADGGRLV--EVOLSGAPWGFTLKGREHGEPLIVTKIEGSKAAAV-DKLLA 70

QY 73 GDEVVHINEVTLSSSRKEAVSLVKGSKYTLRLVVR 107
DB 71 GDEIVGINDIGLSGPRQEAICLVKSGHKTLLVVK 105

RESULT 2
DLG2_RAT STANDARD; PRT; 852 AA.
ID DLG2_RAT AC Q63622; Q62939; P70548;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSYN-110) (SYNAPTIC CHANNEL ASSOCIATED PROTEIN PSD-93) (DISCS, LARGE HOMOLOG 2).
DE DENSITY PROTEIN PSD-93 (DISCS, LARGE HOMOLOG 2).
GN DLG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP MEDLINE=96310881; PubMed=8755482;
RA Kim E., Cho K.O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins.";
RL Neuron 17:103-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96193770; PubMed=8625413;
RA Brennan J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E., Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F., Froehner S.C., Bret D.S.;
RT "Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
RL Cell 84:757-767(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Irie M., Hata Y., Takai Y.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: UA9049; AAB53243.1; .
DR EMBL: U50717; AAC52643.1; .
DR EMBL: U53368; AAB48562.1; .
DR HSSP: Q12959; IPDR.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.

SMART; SM00072; GuKc; 1.
SMART; SM00228; PDZ; 3.
SMART; SM00326; SH3; 1.
PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
PROSITE; PS50106; PDZ; 3.
PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 98 184
FT DOMAIN 193 279
FT DOMAIN 421 501
FT DOMAIN 536 606
FT DOMAIN 662 852
FT DOMAIN 181 182
FT CONFLICT 228 228
FT CONFLICT 326 326
FT CONFLICT 339 339
FT CONFLICT 450 454
FT CONFLICT 464 465
FT CONFLICT 474 474
FT CONFLICT 476 476
FT CONFLICT 478 478
FT CONFLICT 484 486
FT CONFLICT 506 506
FT CONFLICT 569 569
FT CONFLICT 586 586
FT CONFLICT 626 641
FT CONFLICT 639 639
FT CONFLICT 726 726
FT CONFLICT 733 733
FT CONFLICT 749 749
FT CONFLICT 756 756
FT CONFLICT 791 792
FT CONFLICT 794 794
FT SEQUENCE 852 AA; 94934 MW; F8D414A8B9CF5B09 CRC64;

Query Match. 23.1%; Score 126; DB 1; Length 852;
Best Local Similarity 32.7%; Pred. No. 5.3e-05;
Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;

QY 10 KPSATLSNATKQ--RYVILEAPLEGAPWGFTLKGLEHGEPLIIISKVEEGKADTLSS 67
DB 404 QPSVTLQRAISLEGEPRKVVLA---KGTGLGFNVIG-EDGEGIFVFSFLAGGPAD-LS 458

QY 68 SKLAQAGDEVVHINEVTL-SSSRKEAVSLVKGSKYTLRLVVR 107
DB 459 GELQRGDQLSVNGIDLRGASHEQAAALKGAGQTVTIIAQ 499

RESULT 3
DLG2_HUMAN STANDARD; PRT; 870 AA.
ID DLG2_HUMAN AC Q15700;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSYN-110) (DISCS, LARGE HOMOLOG 2).
GN DLG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96310881; PubMed=8755482;
RA Kim E., Cho K.O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins.";
RL Neuron 17:103-113(1996).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS AS WELL AS POTASSIUM CHANNELS.

```

CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32376; AAB04949.1; -.
CC HSSP: Q12959; 1PDR.
CC MIM: 603583; -.
CC InterPro: IPR000619; Guanylate_kin.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001452; SH3.
CC Pfam: PF00625; Guanylate_kin; 1.
CC Pfam: PF00595; PDZ; 3.
CC Pfam: PF00018; SH3; 1.
CC SMART: SM00072; Gukc; 1.
CC SMART: SM00228; PDZ; 3.
CC SMART: SM00326; SH3; 1.
CC SMART: SM00326; SH3; 1.
CC PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
CC PROSITE: PS00852; GUANYLATE_KINASE_2; 1.
CC PROSITE: PS50106; PDZ; 3.
CC PROSITE: PS50002; SH3; 1.
CC SH3 domain; Repeat.
CC SH3 domain; Repeat.
CC PDZ 1.
CC PDZ 2.
CC PDZ 3.
CC PDZ 3.
CC SH3.
CC GUANYLATE KINASE.
CC 870 AA; 97500 MW; 89083BA0619FEF59 CRC64;
SO SEQUENCE

```

Query Match 23.1%; Score 126; DB 1; Length 870;
 Best Local Similarity 32.7%; Pred. No. 5.4e-05;
 Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;

```

OY 10 KRSATINSTATKRG--RYVLEAFLEGAPNPGFTLKGLHEPPLISVEEGKADTUS 67
   :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 404 QPSMTLQRAVSLSEGEPRKRVYLH--KGSGLGFNTVGG-EDGGLVPSFTLAGGPAD-IS 458
   :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 68 SKLQNGDEVHINEVTL-SSSRKAVSLVSGSKYKTLRLVVR 107
   :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 459 GELQRGDQLLSVNGIDLRAASHQAAALAKGAGQVTITIAQ 499
   :||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 4
 PTN3_HUMAN STANDARD; PRT; 913 AA.
 ID PTN3_HUMAN
 AC P26045;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 3 (EC 3.1.3.48)
 DE (PROTEIN-TYROSINE PHOSPHATASE H1) (PTP-H1).
 GN PTPN3 OR PTPH1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP MEDLINE-91296738; PubMed-1648725;
 RA Yang Q., Tonks N.K.;
 RT "Isolation of a cDNA clone encoding a human protein-tyrosine
 RT phosphatase with homology to the cytoskeletal-associated proteins
 RT band 4.1, ezrin, and talin."
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5949-5953(1991).

```

RN [2]
RN SEQUENCE OF 194-896 FROM N.A.
RP TISSUE-colon;
RC MEDLINE-92327504; PubMed-1626183;
RA Altamura Y., Hinoda Y., Itoh F., Takekawa M., Tsujiaki M., Adachi M.,
RA Imai K., Yachi A.;
RT "cDNA cloning of new protein tyrosine phosphatases in the human
RT colon.";
RL Tumour Biol. 13:180-186(1992).
RN [3]
RN SEQUENCE OF 899-913 FROM N.A.
RP MEDLINE-95179278; PubMed-7874267;
RX Ikuta S., Itoh F., Hinoda Y., Toyota M., Makiguchi Y., Imai K.,
RA Yachi A.;
RT "Expression of cytoskeletal-associated protein tyrosine phosphatase
RT PTPH1 mRNA in human hepatocellular carcinoma.";
RL J. Gastroenterol. 29:727-732(1994).
CC -1- FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
CC CYTOSKELETON.
CC -1- CATALYTIC ACTIVITY: PROTEIN-TYROSINE PHOSPHATASE + H(2)O =
CC PROTEIN-TYROSINE + ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M64572; AAA35647.1; -.
CC EMBL: S39392; AAB22439.2; -.
CC EMBL: S76309; AAB33583.1; -.
CC PIR: A41109; A41109.
CC HSSP: P18031; 1PRT.
CC MIM: 176877; -.
CC InterPro: IPR000299; Band_4.1.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR000387; TYR_phosphatase.
CC InterPro: IPR000242; TYR_prot_phphatase.
CC Pfam: PF00373; Band_4.1; 1.
CC Pfam: PF00595; PDZ; 1.
CC Pfam: PF00102; Y-phosphatase; 1.
CC PRINTS: PR00700; PRYPHPTASE.
CC PRINTS: PR00935; BAND41.
CC SMART: SM00295; B41; 1.
CC SMART: SM00228; PDZ; 1.
CC SMART: SM00194; PTPc; 1.
CC PROSITE: PS00660; BAND_41_1; 1.
CC PROSITE: PS00661; BAND_41_2; 1.
CC PROSITE: PS50057; BAND_41_3; 1.
CC PROSITE: PS50106; PDZ; 1.
CC PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
CC PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
CC PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
CC PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
CC Structural protein; cytoskeleton; Hydrolase.
CC KW DOMAIN 84; BAND 4.1-LIKE.
CC FT DOMAIN 510; 582; PDZ.
CC FT DOMAIN 670; 913; PROTEIN-TYROSINE PHOSPHATASE.
CC FT ACT_SITE 842; 842; BY SIMILARITY.
CC SEQUENCE 913 AA; 104029 MW; 29A539ACDE2F1515 CRC64;
SO SEQUENCE

```

Query Match 22.5%; Score 122.5; DB 1; Length 913;
 Best Local Similarity 27.7%; Pred. No. 0.00013;
 Matches 33; Conservative 23; Mismatches 44; Indels 19; Gaps 4;

OY 6 EDPRKPSATLNSNTATKGR-----LYLEAFLEGAPNPGFTLKGLHEPPLIS 53

Db 479 DFHR--VTGGSTEDASQYCDKNDGSDYLVLRITPDEDGKFGNKLKGGVDQKMPLV 536
QY 54 ISKVEEGKADTLSSKLOAGDEVVHNEVTLSS-SRKEAVSLVKGSKYKT-----LRLVVR 107
Db 537 VSRINPSPADTCIPKLNQGDQIVLNGRDISEHSHDQVVMFKAISGSHRELALVIR 595

RESULT 5
PDL1_RAT
ID PDL1_RAT STANDARD; PRT; 327 AA.
AC P52944;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDL1 AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
DE LIM DOMAIN PROTEIN 1) (ELFIN).
GN PDL1 OR CLIM1 OR CLP36.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=96096533; PubMed=8522188;
RA Wang H., Harrison-Shostak D.C., Lemasters J.J., Herman B.;
RT "Cloning of a rat cDNA encoding a novel LIM domain protein with high
RT homology to rat RIL";
RL Gene 165:267-271(1995).
CC -!- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON (BY SIMILARITY).
CC -!- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN HEART, LUNG AND
CC LIVER. MODERATELY IN SPLEEN AND SKELETAL MUSCLE, AND AT EXTREMELY
CC LOW LEVELS (IF AT ALL) IN TESTIS AND BRAIN TISSUES.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U23769; AAA92046.1; -
DR InterPro; IPR001781; LIM.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00412; LIM; 1.
DR Pfam; PF00595; PDZ; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Cytoskeleton; LIM domain; Metal-binding; Zinc.
FT DOMAIN 3 84 PDZ.
FT DOMAIN 258 308 LIM.
SQ SEQUENCE 327 AA; 35525 MW; 4C3B5F579FBDA4D1 CRC64;

Query Match 20.8%; Score 113.5; DB 1; Length 327;
Best Local Similarity 39.5%; Pred. No. 0.00032;
Matches 30; Conservative 11; Mismatches 32; Indels 3; Gaps 2;
QY 32 LEGGAPWGFTLKGLEHCEPLIISKVEEGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90
Db 8 LOGPGWGFRLVGGKDFEOPLAISRTVPGSKA--AIANLCIGDLTAIDGDTSSMTHLE 65

QY 91 AVSLVKGSKYKTLRLVV 106
Db 66 AQNKIKGCVDNMTLV 81

RESULT 6
RIL_RAT
ID RIL_RAT STANDARD; PRT; 330 AA.
AC P36202;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN) (RIT-18) (H-REV18).
GN RIL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=95124724; PubMed=7824279;
RA Kiess M., Scharm B., Aguzzi A., Hajnal A., Klemeniz R.,
RA Schwarte-Waldhoff I., Schaefer R.;
RT "Expression of ril, a novel LIM domain gene, is down-regulated in
RT Hras-transformed cells and restored in phenotypic revertants.";
RL Oncogene 10:61-68(1995).
CC -!- TISSUE SPECIFICITY: DETECTED IN SEVERAL TISSUES, MOST PROMINENT IN
CC BRAIN AND HEART OF ADULTS.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X76454; CAA53992.1; -
DR PIR; S39246; S39246.
DR InterPro; IPR001781; LIM.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00412; LIM; 1.
DR Pfam; PF00595; PDZ; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
DR PROSITE; PS50106; PDZ; 1.
KW LIM domain; Metal-binding; Zinc.
FT DOMAIN 1 84 PDZ.
FT DOMAIN 255 305 LIM.
SQ SEQUENCE 330 AA; 35521 MW; 84BEB1E629ADA584 CRC64;

Query Match 20.6%; Score 112.5; DB 1; Length 330;
Best Local Similarity 40.8%; Pred. No. 0.00041;
Matches 31; Conservative 9; Mismatches 33; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLEHCEPLIISKVEEGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90
Db 7 LRGSFPMGFRVLVGGKDFEAPLISRVHAGSKA--ALAAALCPGDSIQAINGESTELMTHLE 64
QY 91 AVSLVKGSKYKTLRLVV 106
Db 65 AQNKIKGCHDHLTV 80

```

RESULT 7
ID DGL1_RAT STANDARD; PRT; 911 AA.
AC 062696;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
DE LARGE HOMOLOG 1).
GN DGL1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=95198112; PubMed=7891172;
RA Mueller B.M., Kistner U., Ven R.W., Cases-Langhoff C., Becker B.,
RA Gundelfinger E.D., Garner C.C.;
RT "Molecular characterization and spatial distribution of SAP97, a
RT novel presynaptic protein homologous to SAP90 and the Drosophila
RT discs-large tumor suppressor protein."
RL J. Neurosci. 15:2354-2366(1995).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN
CC CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO
CC THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANS BETWEEN ADJACENT
CC CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
CC -1- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY
CC SYNAPSES. IN OTHER BRAIN REGIONS, SAP97 IS FOUND IN AND ALONG
CC BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS,
CC BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A
CC VARIETY OF EPITHELIAL CELLS.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14950; AAA79976.1; -.
DR HSP: Q12959; 1PDR.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00072; Gukc; 1.
DR SMART: SM00228; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE: PS00552; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS0106; PDZ; 3.
DR PROSITE: PS0002; SH3; 1.
DR SH3 domain; Repeat.
FT DOMAIN 224 310 PDZ 1.
FT DOMAIN 318 404 PDZ 2.
FT DOMAIN 465 545 PDZ 3.
FT DOMAIN 580 650 SH3.
FT DOMAIN 721 911 GUANYLATE_KINASE.
FT DOMAIN 527 530 POLY-ALA.
SEQUENCE 911 AA; 100570 MW; 18CEBD31DDCAFB8 CRC64;

```

```

Query Match 20.6%; Score 112.5; DB 1; Length 911;
Best Local Similarity 27.9%; Pred. No. 0.0013;
Matches 34; Conservative 27; Mismatches 42; Indels 19; Gaps 5;

QY 3 RTTEDEHFKPSATINSNTATGRYIL-ENFL-----EGAPMGFTLKGL 46
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 424 QSDVNHVSPSSYLGQVPASPARSPISKAVLGDEITREPRKYVLHKGSTGLFNIVGG- 482
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 47 EHGEPLIISVEBGKADITSSKLQAGDEVHINEVTL-SSSRKENVSLYKSGYKTLRLV 105
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 483 EDGEGIFISFIAGGPAID-LSGELRKGDRIISVNSVDLRAASHQAAALNAGAVTV 541
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 106 VR 107
DB 542 AQ 543

RESULT 8
ID DGL1_HUMAN STANDARD; PRT; 904 AA.
AC Q12959; Q12958;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
DE LARGE HOMOLOG 1).
GN DGL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=95024052; PubMed=7937897;
RA Line R.A., Marfatia S.M., Branton D., Chishti A.H.;
RA "Cloning and characterization of hdlg: the human homologue of the
RA Drosophila discs large tumor suppressor binds to protein 4.1."
RL Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
CC [2]
CC X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
CC MEDLINE=9638231; PubMed=8757139;
CC Capral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
CC Marfatia S.M., Chishti A.H., Liddington R.C.;
CC "Crystal structure of a PDZ domain."
CC Nature 382:649-652(1996).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. ASSOCIATES WITH PROTEIN 4.1.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U13897; AAA50599.1; -.
DR EMBL: U13896; AAA50598.1; -.
DR PDB: 1PDR; 23-JUL-97.
DR MIM: 601014; -.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00072; Gukc; 1.

```

```
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain; Repeat; Alternative splicing; 3D-structure.
KW DOMAIN 224 310 PDZ 1.
FT DOMAIN 319 405 PDZ 2.
FT DOMAIN 466 546 PDZ 3.
FT DOMAIN 581 651 SH3.
FT DOMAIN 714 904 GUANYLATE_KINASE.
FT VARSPLIC 669 680 EIPDDMSKGLK -> QSFNDKRRKNLFSRKFPEYKKNQDS
FT EQETSDAQDQ (IN ISOFORM 2).
FT SEQUENCE 904 AA; 100354 MW; B78798D6B0920D4 CRC64;

Query Match 20.4%; Score 111; DB 1; Length 904;
Best Local Similarity 30.3%; Pred. No. 0.0018;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

QY 7 DFH-KFSATLNSWTATKGRYIVL-EAFL-----EGGAPMGFTLKGGLHG 49
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 428 DNVSPSFLGQTPASPARSPVSKAVLGDDITREPRKVLHRTGTGFGFNIVGG-EDG 486
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 50 EPLIISKVEGGKADTLSSKLQAGDEWHINEVTL-SSSRKEAVSLVKGYSKTLRLVVR 107
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 487 EGIFISFILAGGPAD-LSGELRKGDRIISVNSVDLRAASHEQAQAALKNAGQAVTVAQ 544
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
RIL_CHICK
ID RIL_CHICK STANDARD; PRT; 330 AA.
AC QSPW72; 2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN).
GN RIL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn;
RX MEDLINE-20379358; PubMed=10918612;
RA Fu S.-L., Waha A., Vogt P.K.;
RT "Identification and characterization of genes upregulated in cells transformed by v-Jun.";
RL Oncogene 19:3537-3545(2000).
CC -!- FUNCTION: NEGATIVE REGULATOR OF CELL GROWTH.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC IONS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; AF167295; AAD46655.1;
CC InterPro; IPR001781; LIM.
CC InterPro; IPR001478; PDZ.
CC Pfam; PF00412; LIM; 1.
CC Pfam; PF00593; PDZ; 1.
CC ProDom; PD000094; LIM; 1.
CC SMART; SM00132; LIM; 1.
CC SMART; SM00228; PDZ; 1.
```

```
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS00023; LIM_DOMAIN_2; 1.
DR PROSITE; PS50106; PDZ; 1.
KW LIM domain; Metal-binding; Zinc.
FT DOMAIN 8 84 PDZ.
FT DOMAIN 255 305 LIM.
SQ SEQUENCE 330 AA; 35838 MW; 2C0E0B404301E404 CRC64;

Query Match 19.9%; Score 108.5; DB 1; Length 330;
Best Local Similarity 38.2%; Pred. No. 0.001;
Matches 2; Conservative 12; Mismatches 32; Indels 3; Gaps 2;

QY 32 LEGAPMGFTLKGLEHGEPLIISKVEGGKADTLSSKLQAGDEWHINEVTLSSSRKE 90
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7 LRGPSPMGFRLVGGKDFSTPLTISRNPGSKA--ALANLCPGDIILAINGESTAMTHLE 64
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 91 AVSLVKGYSKTLRLVV 106
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 AQNKIKACVEQLLSV 80
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
DLG3_MOUSE
ID DLG3_MOUSE STANDARD; PRT; 849 AA.
AC P70175;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102) (DISCS, LARGE HOMOLOG 3).
DE DLG3 OR DLGH3.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Kohmura N., Makino S., Yagi T.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR SUBUNIT NR2B.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS A GUANYLATE_KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; D87117; BAAL3249.1;
CC HSP; Q12959; IPDR.
CC MGD; MG1:188986; Dlg3.
CC InterPro; IPR000619; Guanylate_kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00625; Guanylate_kin; 1.
CC Pfam; PF00595; PDZ; 3.
CC Pfam; PF00018; SH3; 1.
CC SMART; SM00072; GUKC; 1.
CC SMART; SM00228; PDZ; 3.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
CC PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
CC PROSITE; PS0106; PDZ; 3.
CC PROSITE; PS50002; SH3; 1.
CC SH3 domain; Repeat.
```

FT DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE KINASE.
SO SEQUENCE 849 AA; 93482 MM; EF3EF2D/51538EE CRC64;

Query Match 19.5%; Score 106.5; DB 1; Length 849;
Best Local Similarity 32.4%; Pred. No. 0.0048; Indels 11; Gaps 6;
Matches 34; Conservative 23; Mismatches 37;

QY 10 KPSATLNSNTATKGRYIYLEAFLE-GGAPMGFTLKGLH-GEPLIISKVEEGKA 63
DB 133 EPSLSVN---GSDGMFYEEIYLERNGSLGFSIAGGIDNPVDDPGIFTKIIIPGGA 189
DB 64 DTLSSKIQAGDEVVHINEVTLSS-SRKEAVSLVSGSYKTLVLVR 107
DB 190 -AMGRLGVNDCVLRVNEVDSEVYHSRAVEALKRAGVRLVVR 233

RESULT 11
ID DLG3_RAT STANDARD; PRT; 849 AA.
AC 062936; P70547; (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102) (PSD-95/SAP90 RELATED PROTEIN 1) (DISCS, LARGE HOMOLOG 3).
GN DLG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid-10116;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE-Brain;
RX MEDLINE-96374358; PubMed-8780649;
RA Mueller B.M., Kistner U., Kindler S., Kuhlendahl S., Penster S.D., Lau L.-F., Veb R.W., Huganir R.L., Gundelfinger E.D., Garner C.C.,
RA "SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes in vivo."
RT Neuron 17:255-265(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RA Title M., Hata Y., Takai Y.;
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR SUBUNIT NR2B.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: U50147; AAA93031.1; -
DR EMBL: U53367; AAA84561.1; -
DR HSSP: Q12959; IPDR.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.

DR SMART: SM00072; GUKC; 1.
DR SMART: SM00228; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
KM SH3 domain; Repeat; Alternative splicing.
FT DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE KINASE.
FT VARSPLIC 627 640 MISSING (IN SHORT ISOFORM).
SO SEQUENCE 849 AA; 93539 MM; 34DA9C46C7BB96DB CRC64;

Query Match 19.5%; Score 106.5; DB 1; Length 849;
Best Local Similarity 32.4%; Pred. No. 0.0048; Indels 11; Gaps 6;
Matches 34; Conservative 23; Mismatches 37;

QY 10 KPSATLNSNTATKGRYIYLEAFLE-GGAPMGFTLKGLH-GEPLIISKVEEGKA 63
DB 133 EPSLSVN---GSDGMFYEEIYLERNGSLGFSIAGGIDNPVDDPGIFTKIIIPGGA 189
DB 64 DTLSSKIQAGDEVVHINEVTLSS-SRKEAVSLVSGSYKTLVLVR 107
DB 190 -AMGRLGVNDCVLRVNEVDSEVYHSRAVEALKRAGVRLVVR 233

RESULT 12
ID PDL1_HUMAN STANDARD; PRT; 329 AA.
AC 000151; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL LIM DOMAIN PROTEIN 1) (ELFIN).
GN PDLIM1 OR CLIM1 OR CLP36.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid-9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Heart.
RX MEDLINE-99145108; PubMed-10022510;
RA Kotaka M., Ngai S.W., Garcia-Barcelo M., Tsui S.K.W., Fung K.P., Lee C.Y., Waye M.M.Y.;
RA "Characterization of the human 36-kDa carboxyl terminal LIM domain protein (hCLIM1)."
RT J. Cell. Biochem. 72:279-285(1999).
RN [2]
RP INTERACTION WITH ALPHA-ACTININ 2.
RX MEDLINE-2030365; PubMed-10861853;
RA Kotaka M., Kostin S., Ngai S., Chan K., Lau Y., Lee S.M., Li H.Y., Ng E.K., Schaper J., Tsui S.K.W., Fung K.P., Lee C.Y., Waye M.M.Y.;
RA "Interaction of hCLIM1, an enigma family protein, with alpha-actinin 2."
RT J. Cell. Biochem. 78:558-565(2000).
RN [3]
RP INTERACTION WITH ALPHA-ACTININS 1 AND 4.
RX MEDLINE-20219155; PubMed-10753915;
RA Vallianus T., Luukko K., Mäkelä T.P.;
RA "CLP-36 PDZ-LIM protein associates with nonmuscle alpha-actinin-1 and alpha-actinin-4."
RT J. Biol. Chem. 275:11100-11105(2000).
CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT BRINGS OTHER PROTEINS TO THE CYTOSKELETON.
CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN THE HEART AND SKELETAL


```
CC MUSCLE, MODERATELY EXPRESSED IN THE SPLEEN, SMALL INTESTINE,  
CC COLON, PLACENTA, AND LUNG. A LOWER LEVEL EXPRESSION IS SEEN IN  
CC LIVER, THYMUS, KIDNEY, PROSTATE AND PANCREAS AND IS NOT FOUND IN  
CC THE BRAIN, TESTIS, OVARY, AND PERIPHERAL BLOOD LEUCOCYTES.  
CC -|- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC  
CC IONS.  
CC -|- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U90878; AAC05580.1; -  
CC MIM; 605900; -  
CC InterPro; IPR001781; LIM.  
CC InterPro; IPR001478; PDZ.  
CC Pfam; PF00412; LIM; 1.  
CC Pfam; PF00595; PDZ; 1.  
CC ProDom; PD000094; LIM; 1.  
CC SMART; SM00132; LIM; 1.  
CC SMART; SM00228; PDZ; 1.  
CC PROSITE; PS00478; LIM_DOMAIN_1; FALSE_NEG.  
CC PROSITE; PS00023; LIM_DOMAIN_2; 1.  
CC PROSITE; PS00106; PDZ; 1.  
CC Cytoskeleton; LIM domain; Metal-binding; Zinc.  
CC FT DOMAIN 3 84 PDZ.  
CC FT DOMAIN 260 310 LIM.  
CC SEQUENCE 329 AA; 36171 MW; 6520F1A932CFA312 CRC64;  
  
Query Match 19.2%; Score 104.5; DB 1; Length 329;  
Best Local Similarity 38.2%; Pred. No. 0.0026;  
Matches 29; Conservative 11; Mismatches 33; Indels 3; Gaps 2;  
  
QY 32 LEGAPMGFLKGLGHEPLIISKVEGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90  
Db 8 LQGGPGWFLVGRKDFEQPLAISRYTPGSKA--ALANLCIGDVITADIGTNTSNTHLE 65  
QY 91 AVSLVKGSYKTLRLV 106  
Db 66 AQNRKICTDNLITV 81  
  
RESULT 13  
DLG4_MOUSE ID DLG4_MOUSE STANDARD; PRT; 724 AA.  
AC Q62108;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PRESYNAPTIC DENSITY PROTEIN 95 (PSD-95) (PRESYNAPTIC PROTEIN SAP90)  
DE (SYNAPSE-ASSOCIATED PROTEIN 90) (DISCS, LARGE HOMOLOG 4).  
GN DLG4 OR DLG4 OR PSD95.  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DKA/2; TISSUE=Brain;  
RA Kohmura N., Yagi T.;  
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
CC -|- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR  
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY  
CC ON THE PRESYNAPTIC SIDE (BY SIMILARITY).  
CC -|- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.  
CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -|- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
```

```
CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; D50621; BAA09297.1; -  
CC HSSP; Q12959; LPDR.  
CC MGD; MGI:1277959; Digh4.  
CC InterPro; IPR000619; Guanylate_kin.  
CC InterPro; IPR001478; PDZ.  
CC InterPro; IPR001452; SH3.  
CC Pfam; PF00625; Guanylate_kin; 1.  
CC Pfam; PF00595; PDZ; 3.  
CC Pfam; PF00018; SH3; 1.  
CC SMART; SM00072; GUKC; 1.  
CC SMART; SM00228; PDZ; 3.  
CC SMART; SM00326; SH3; 1.  
CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.  
CC PROSITE; PS00552; GUANYLATE_KINASE_2; 1.  
CC PROSITE; PS0106; PDZ; 3.  
CC PROSITE; PS00002; SH3; 1.  
CC SH3 domain; Repeat.  
CC FT DOMAIN 65 151 PDZ 1.  
CC FT DOMAIN 160 246 PDZ 2.  
CC FT DOMAIN 313 393 PDZ 3.  
CC FT DOMAIN 428 498 SH3.  
CC FT DOMAIN 534 724 GUANYLATE KINASE.  
CC SEQUENCE 724 AA; 80472 MW; 7E9F99E1FF90BA CRC64;  
  
Query Match 19.2%; Score 104.5; DB 1; Length 724;  
Best Local Similarity 34.7%; Pred. No. 0.0063;  
Matches 26; Conservative 20; Mismatches 26; Indels 3; Gaps 3;  
  
QY 34 GGAPMGFTLKGLGHEPLIISKVEGKADTLSSKLOAGDEVVHIN-EVTL-SSSRKEAV 92  
Db 319 GSTGLGFNIVGG-EDGEIGFISFILAGPAD-LSGELRGDQILSYNGVDLRNASHEQAA 376  
QY 93 SLVKGSYKTLRLV 107  
Db 377 IALKNAGQTVTIAQ 391  
  
RESULT 14  
DLG4_RAT ID DLG4_RAT STANDARD; PRT; 724 AA.  
AC P31016; P97631;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PRESYNAPTIC DENSITY PROTEIN 95 (PSD-95) (PRESYNAPTIC PROTEIN SAP90)  
DE (SYNAPSE-ASSOCIATED PROTEIN 90) (DISCS, LARGE HOMOLOG 4).  
GN DLG4 OR PSD95.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;  
RX MEDLINE=93040233; PubMed=1419001;  
RA Cho K.-O., Hunt C.A., Kennedy M.B.;  
RT "The rat brain postsynaptic density fraction contains a homolog of  
RT the Drosophila discs-large tumor suppressor protein.";  
RL Neuron 9:929-942(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
```


RX MEDLINE-93186749; PubMed-7680343; Cases-Langhoff C., Garner A.M.,
 RA Kistner U., Menzel B.M., Voh R.W., Gundelfinger E.D., Garner C.C.,
 RA Appeltner U., Voss B., Gundelfinger E.D., Garner C.C.,
 RT *SAP90, a rat presynaptic protein related to the product of the
 RT Drosophila tumor suppressor gene dlg-A.";
 RL J. Biol. Chem. 268:4580-4583(1993).
 RN (3)
 RP SEQUENCE OF 566-625 FROM N.A.
 RC STRAIN-WISTAR KYOTO; TISSUE=Vascular smooth muscle;
 RA Adams L.D., Werry I., Schwartz S.M.;
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS) OF 302-402.
 RX MEDLINE-96270509; PubMed-8674113;
 RA Doyle D.A., Lee A., Lewis J., Kim E., Sheng M., Mackinnon R.;
 RT *Crystal structures of a complexed and peptide-free membrane protein-
 RT binding domain: molecular basis of peptide recognition by PDZ.";
 RL Cell 85:1067-1076(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY
 CC ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
 CC POSTSYNAPTIC).
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M96853; AAA41971.1; -
 DR EMBL; X66474; CAA47103.1; -
 DR EMBL; U77090; AAB38270.1; -
 DR PIR; S26407; S26407.
 DR PIR; JH0800; JH0800.
 DR PDB; 1BR9; 21-OCT-98.
 DR PDB; 1BR9; 21-OCT-98.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR SMART; SM00018; SH3; 1.
 DR SMART; SM00072; GUKK; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50106; PDZ; 3.
 DR PROSITE; PS50002; SH3; 1.
 KW SH3 domain; Repeat; 3D-structure.
 FT DOMAIN 65 151 PDZ 1.
 FT DOMAIN 160 246 PDZ 2.
 FT DOMAIN 313 393 PDZ 3.
 FT DOMAIN 428 498 SH3.
 FT DOMAIN 534 724 GUANYLATE_KINASE.
 FT CONFLICT 61 78 M -> L (IN REF. 2).
 FT CONFLICT 76 78 S -> T (IN REF. 2).
 FT CONFLICT 177 182 GVGNQH -> ALGTSI (IN REF. 2).
 FT CONFLICT 200 200 A -> G (IN REF. 2).
 FT CONFLICT 254 254 S -> T (IN REF. 2).
 FT CONFLICT 539 555 ILGPTKRDANDDLISER -> ISLDPPKTVPTMIFSPSS
 FT CONFLICT (IN REF. 2).
 FT CONFLICT 623 625 GKX -> RDO (IN REF. 3).
 FT SEQUENCE 724 AA; 80465 MW; 7922D4B8E0F9AD5 CRC64;

Query Match 19.2%; Score 104.5; DB 1; Length 724;
 Best Local Similarity 34.7%; Pred. No. 0.0063;
 Matches 26; Conservative 20; Mismatches 26; Indels 3; Gaps 3;
 OY 34 GGAWGFTLKGLEHPELISKEVEGGKADTLSSKLQADDEVVHNEVTL-SSSRKRAY 92
 Db 319 GSTLGFNIVGG-EDGEIFISFILAGPAD-LSGEIRKBDQILSVNGVDLRNASHBOAA 376
 OY 93 SLVGSYKTLRLVLR 107
 Db 377 IALKNAGOTVRIIAQ 391
 RESULT 15
 DLG4_HUMAN STANDARD; PRI; 767 AA.
 ID DLG4_HUMAN
 AC P78352; Q92941; Q90RK8;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC DENSITY PROTEIN 95 (PSD-95) (DISCS, LARGE HOMOLOG 4).
 GN DLG4 OR PSD95.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE-97432822; PubMed-9286702;
 RA Stathakis D.G., Hoover K.B., You Z., Bryant P.J.;
 RT "Human postsynaptic density-95 (PSD95): location of the gene (DLG4)
 RT and possible function in nonneural as well as in neural tissues.";
 RL Genomics 44:71-82(1997).
 RN [2]
 RP REVISIONS.
 RC TISSUE=Mammary gland;
 RA Stathakis D.G., Hoover K.B., You Z., Bryant P.J.;
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20047407; PubMed-10582582;
 RA Stathakis D.G., Udar N., Sandgren O., Andreasson S., Bryant P.J.,
 RA Small K., Forsman-Semb K.;
 RT "Genomic organization of human DLG4, the gene encoding postsynaptic
 RT density 95.";
 RL J. Neurochem. 73:2250-2265(1999).
 RN [4]
 RP SEQUENCE OF 81-401 FROM N.A.
 RC TISSUE=Brain;
 RA Brenman J.E., Bredt D.S., Parkinson J.F., Manzana W.P., McClary J.A.;
 RL Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY
 CC ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
 CC POSTSYNAPTIC).
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U83192; AAC52113.1; -

```
DR EMBL; AF156495; AAD56173.1; -.
DR EMBL; U68138; AAB07736.1; -.
DR HSP; Q12959; IPDR.
DR MIM; 602897; -.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR01478; PDZ.
DR InterPro; IPR01452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS0052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS00002; SH3; 1.
DR SH3 domain; Repeat.
KW DOMAIN 108 194 PDZ 1.
FT DOMAIN 203 289 PDZ 2.
FT DOMAIN 356 436 PDZ 3.
FT DOMAIN 471 541 SH3.
FT DOMAIN 577 767 GUANYLATE_KINASE.
FT CONFLICT 46 46 E -> V (IN REF. 3).
FT CONFLICT 81 83 VIV -> EFR (IN REF. 4).
FT CONFLICT 399 401 GDO -> AGI (IN REF. 4).
SQ SEQUENCE 767 AA; 85429 MW; BE1019159E65B2D8 CRC64;

Query Match 19.2%; Score 104.5; DB 1; Length 767;
Best Local Similarity 34.7%; Pred. No. 0.0068;
Matches 26; Conservative 20; Mismatches 26; Indels 3; Gaps 3;

QY 34 GGAPWGFTIKGGLGEPLIISKVEEGKADTLSSKLOAGDEVVHINEVTL-SSSRKEAV 92
DB 362 GSTGLGFTNIVGG-EDGEGIFISFILAGGPAD-LSGELRKGDQILSVNGVDLRNASHEQAA 419
QY 93 SLVKGSYKTLRLVVR 107
DB 420 IALKNAGQTVIIAQ 434
```

Search completed: January 31, 2002, 21:06:28
Job time: 51 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:56:12 ; Search time 56.12 Seconds
(without alignments)
142.550 Million cell updates/sec

Title:	US-09-641-831-4
Perfect score:	545
Sequence:	1 MMRTEDPHKPSATLSNTA.....KEAVSLVKGSYKTLRLVRS 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      522463 segs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

1:	/SID52/gcgdata/genseq/genseqp/A11980.DAT.*
2:	/SID52/gcgdata/genseq/genseqp/A11981.DAT.*
3:	/SID52/gcgdata/genseq/genseqp/A11982.DAT.*
4:	/SID52/gcgdata/genseq/genseqp/A11983.DAT.*
5:	/SID52/gcgdata/genseq/genseqp/A11984.DAT.*
6:	/SID52/gcgdata/genseq/genseqp/A11985.DAT.*
7:	/SID52/gcgdata/genseq/genseqp/A11986.DAT.*
8:	/SID52/gcgdata/genseq/genseqp/A11987.DAT.*
9:	/SID52/gcgdata/genseq/genseqp/A11988.DAT.*
10:	/SID52/gcgdata/genseq/genseqp/A11989.DAT.*
11:	/SID52/gcgdata/genseq/genseqp/A11990.DAT.*
12:	/SID52/gcgdata/genseq/genseqp/A11991.DAT.*
13:	/SID52/gcgdata/genseq/genseqp/A11992.DAT.*
14:	/SID52/gcgdata/genseq/genseqp/A11993.DAT.*
15:	/SID52/gcgdata/genseq/genseqp/A11994.DAT.*
16:	/SID52/gcgdata/genseq/genseqp/A11995.DAT.*
17:	/SID52/gcgdata/genseq/genseqp/A11996.DAT.*
18:	/SID52/gcgdata/genseq/genseqp/A11997.DAT.*
19:	/SID52/gcgdata/genseq/genseqp/A11998.DAT.*
20:	/SID52/gcgdata/genseq/genseqp/A11999.DAT.*
21:	/SID52/gcgdata/genseq/genseqp/A12000.DAT.*
22:	/SID52/gcgdata/genseq/genseqp/A12001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	545	100.0	108	22	AAU00032	Novel human prote
2	545	100.0	180	22	AAU00031	Novel human prote
3	541	99.3	133	22	AAU00033	Novel human prote
4	148.5	23.6	150	21	AAU43364	Human OREF ORF128
5	124.5	22.8	207	22	AAE03655	Human extracellular
6	122.5	22.5	406	19	AAW70326	Secreted protein D
7	122.5	22.5	913	18	AAI12532	Protein tyrosine P
8	122.5	22.5	913	20	AAI25156	Human p191 protein
9	122	22.4	481	20	AAW07095	Colon cancer assoc
10	122	22.4	481	20	AAW07098	Colon cancer assoc
11	122	22.4	521	20	AAW07099	Colon cancer assoc

12	122	22.4	652	20	AAV07094	Colo1 cancer assoc
13	122	22.4	652	20	AAV07050	Renal cancer assoc
14	117.5	21.6	86	22	AABS5880	PDZ encoded domain
15	117.5	21.6	86	22	AABS7631	Human LIM protein
16	117.5	21.6	86	22	AABS8041	Human LIM protein
17	117	21.5	37	22	AAAM2879	Peptide #6912 enco
18	116.5	21.4	127	20	AAV12510	Human 5' EST seque
19	115.5	21.2	197	22	AAAG3600	Human protein sequ
20	113.5	20.6	113	21	AAAG02039	Human secreted pro
21	113.5	20.6	233	20	AAV29978	Human cell junctio
22	112.5	20.6	323	22	AAV07127	Human cell junctio
23	112.5	20.6	329	19	AAAM58336	Human protein sequ
24	112.5	20.6	352	22	AAAG54542	Human PDZ protein
25	112	20.6	181	22	AAAG54378	Human PDZ protein
26	109.5	20.1	316	19	AAAM58397	Human protein HLIM-
27	109.5	20.1	339	21	AAAB56930	Human prostate can
28	108	19.8	97	22	AAAB55836	PDZ encoded domain
29	108	19.8	97	22	AAAB57627	Tax interaction pr
30	108	19.8	97	22	AAAB58037	Tax interaction pr
31	106.5	19.5	95	21	AAAG02034	Human secreted pro
32	106.5	19.5	153	21	AAAB36475	Human LMP-3 (HLMP-
33	106.5	19.5	223	20	AAAB97845	Human LMP-1 amino
34	106.5	19.5	423	21	AAAB36473	Human LMP-2 (HLMP-
35	106.5	19.5	423	21	AAAB36474	Rat bone mineralis
36	106.5	19.5	457	20	AAAB97843	Human bone mineral
37	106.5	19.5	457	20	AAAB97844	Human LIM mineralis
38	106.5	19.5	457	21	AAAB36471	Rat LIM mineralisa
39	106.5	19.5	457	21	AAAB36472	Human LIM minerali
40	104.5	19.2	80	21	AAAB52331	Protein containing
41	104.5	19.2	344	22	AAAB57623	PDZ encoded domain
42	104.5	19.2	344	22	AAAB57623	Human post-synapti
43	104.5	19.2	344	22	AAAB58035	Human post-synapti
44	104.5	19.2	454	17	AAAB8904	Human insulin rece
45	104.5	19.2	882	18	AAAB3662	Partial PSD-93 p

ALIGNMENTS

RESULT	1
ID	AAU00032 standard; Protein; 108 AA.
XX	
XX	
AC	AAU00032;
XX	
DT	11-MAY-2001 (first entry)
XX	
DE	Novel human protein, NHP#2.
XX	
KW	Human; novel human protein; NHP#2; gene therapy; drug screening;
XX	
OS	obesity; high blood pressure.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Misc.feature
FT	1 Location/Qualifiers
FT	/note="Absent if the true start codon is
FT	nucleotides 4-6 of sequence in AA500040, the authors
FT	are uncertain which is the true start codon"
XX	
PN	W0200114422-A1.
XX	
PD	01-MAR-2001.
XX	
PF	18-AUG-2000; 2000WO-US22815.
XX	
PR	24-AUG-1999; 99US-0150511.
XX	
PA	(LEXI-) LEXICON GENETICS INC.
XX	
PI	Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
XX	
DR	WPI; 2001-218430/22.

PT	Novel human polynucleotides isolated from human mammary gland cDNA
PT	library, encodes novel human proteins which are useful in diagnosis,
PT	drug screening, clinical trial monitoring or treating behavioural
PT	disorders -
XX	
XX	Claim 2; Page 24-25; 29pp; English.
XX	
CC	The sequence is a novel human protein, NHP#1, which shares
CC	structural motifs with human APXL protein. Nucleotide constructs
CC	encoding functional NHPs are used in gene therapy approaches for the
CC	modulation of NHP expression. NHP oligonucleotides can be used as
CC	hybridisation probes for screening libraries and assessing NHP gene
CC	expression patterns. Also labeled NHP nucleotide probes can be used to
CC	screen a human genomic library. The NHP nucleotide sequences are also
CC	useful in drug screening techniques for treating symptomatic or
CC	phenotypic manifestations of perturbing the normal function of NHP in the
CC	body. Examples of such manifestations may include obesity and high blood
CC	pressure.
XX	
XX	Sequence 190 AA;
XX	
XX	Sequence 190 AA;

```

Query Match      100.0%; Score 545; DB 22; Length 190;
Best Local Similarity 100.0%; Pred. No. 2.2e-59;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMRTEDEHKPSATLNSNTAKGRYIYLEAFLEGGAPWGLKGGLEHGEPLIISKVEEG 60
      |||||
Db 1 mmrtdedhkpsatlnsntatgryyleafleggapwftlkgglehgepliskveeg 60
      |||||

Qy 61 GKADTLSSKLQAGDEWVHINEVTLSSRKEAVSLVKGYSKTLRLVRS 108
      |||||
Db 61 gkadtlssklqagdewvhinevtlssrkeavslvkgysktrlrvrs 108
      |||||

```

RESULT	3
AAU000033	AAU000033 standard; Protein; 133 AA.
XX	XX
XX	XX
AC	AAU000033;
XX	XX
DT	11-MAY-2001 (first entry)
XX	XX
DE	Novel human protein, NHP#3.
XX	XX
XX	Human; novel human protein; NHP#3; gene therapy; drug screening;
KW	obesity; high blood pressure.
KW	

PN	WO200114422-A1.	
XX		
XX	01-MAR-2001.	
XX		
XX	18-AUG-2000; 2000WO-US22815.	
XX		
XX	24-AUG-1999; 99US-0150511.	
XX		
XX	(LEXI-) LEXICON GENETICS INC.	
XX		
PI	Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;	
XX		
DR	WPI: 2001-218430/22.	
DR	N-PSDB; AAS00041.	
XX		
PT	Novel human polynucleotides isolated from human mammary gland cDNA	
PT	library, encodes novel human proteins which are useful in diagnosis,	
PT		

PT		drug screening, clinical trial monitoring or treating behavioural disorders -
PS	Claim 6; Page 25-26; 29pp;	English.
XX	The sequence is a novel human protein, NHP#3, which shares structural motifs with human APXL protein. Nucleotide constructs encoding functional NHPs are used in gene therapy approaches for the modulation of NHP expression. NHP oligonucleotides can be used as hybridisation probes for screening libraries and assessing NHP gene expression patterns. Also, labeled NHP nucleotide sequences can be used to screen a human genomic library. The NHP nucleotide sequences are also useful in drug screening techniques for treating symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. Examples of such manifestations may include obesity and high blood pressure.	
SQ	Sequence	133 AA;
XX		
OY	Query Match	99.3%; Score 541; DB 22; Length 133; Best Local Similarity 100.0%; Pred. No. 4,2e+59; Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Dd	1 MMRTEDFHKPSATINSNTATGRIYYEALTEGAAPWGTIKGLHCEPILIIKVVEEG 1 mrrtedfhhkpsatinsnatkgriyyleaifleggapwgtikglghcepiliskveeg 60	
Oy	61 GKADPLSSKLGDGEVWHINEVTLISSREAVSLVKSGSKTRLIVLR 61 gkadplssklgdgevwihnevltlissrkeavslvksysktlrilivr 107	
Dd	61 GKADPLSSKLGDGEVWHINEVTLISSREAVSLVKSGSKTRLIVLR 61 gkadplssklgdgevwihnevltlissrkeavslvksysktlrilivr 107	
RESULT 4	AAB43364	
ID	AAB43364 standard; Protein; 150 AA.	
XX	AAB43364;	
AC AC		
DT DT	08-FEB-2001 (first entry)	
XX XX DE	Human ORFX ORFJ128 polypeptide sequence SEQ ID NO:6256.	
XX	Human; open reading frame; ORF; detection; cytosolic; hepatotropic; vulerary; antipsoaristic; antiParkinsonian; neurotropic; neuroprotective; immunosuppressant; osteopathic; antiarthritic; vasodilator; cardiatic; immunoconulant; thrombolitic; coagulant; antididiabetic; hypotesisive; dermatological; immunosuppressive; antiflammatory; antiviral; antibacterial; antifungal; antineumalic; antilyroid; antianeemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; diabetes mellitus; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antinflammatory disease; coagulation; thrombosis; contraceptive.	
KM KM KW KK KS KX OS PN PP PX PR PS PT PA XX	Homo sapiens. WO200058473-A2. 05-OCT-2000. 31-MAR-2000; 200OMO-US08621. 31-MAR-1999; 99US-0127607. PR 02-APR-1999; 99US-0127636. PP 05-APR-1999; 99US-0127728. PX 30-MAR-2000; 200OVS-0540763. (CUBRA-) CUBAGEN CORP.	

[illegible]

XX FH Key Location/Qualifiers
FT Domain 91..171
FT /note= "SH3 domain"
FT Domain 93..174
FT /note= "PDZ domain"
XX WO200142285-A2.
XX 14-JUN-2001.
XX 05-DEC-2000; 2000WO-US32990.
XX 10-DEC-1999; 99US-0172852.
XX 16-DEC-1999; 99US-0172354.
XX (INCY-) INCYTE GENOMICS INC.
XX Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
XX Baughn MR, Lu DAM, Shah P, Au-Young J;
XX WPI: 2001-381632/40.
XX N-PSDB; AAD08063.
XX New human extracellular matrix and cell adhesion molecules and
XX polynucleotide sequences encoding them, useful for diagnosis,
XX prevention, treatment of genetic, autoimmune and cell proliferative
XX disorders
XX Claim 1; Page 111-112; 135pp; English.
XX The present sequence is a human extracellular matrix and cell
XX adhesion molecule (XNAD). The XNAD is used for screening a compound for
XX effectiveness as an agonist or antagonist of XNAD. The identified agonist
XX or antagonist are used for treating a disease or condition associated
XX with decreased or increased expression of functional XNAD. The
XX polynucleotides encoding XNAD are useful in somatic or germline gene
XX therapy to correct a genetic deficiency, to express a conditionally
XX lethal gene product and to express a protein which affords protection
XX against intracellular parasites and also for diagnosis of disorders
XX associated with expression of XNAD. They are also used for generating
XX hybridisation probes useful in mapping the naturally occurring genomic
XX sequences and to create knock in humanised animals (pigs) or transgenic
XX animals (mice or rats) to model human diseases. Oligonucleotide or longer
XX fragments derived from the polynucleotide sequences may be used as
XX elements on a microarray. Antibodies which specifically bind XNAD may be
XX used for the diagnosis of disorders associated with the expression of
XX XNAD, or in assays to monitor patients being treated with XNAD. Diseases
XX diagnosed, prevented or treated include genetic disorders such as
XX adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
XX disease, myotonic dystrophy, sickle cell anaemia, thalasassaemia,
XX autoimmune/inflammatory disorders such as acquired immune deficiency
XX syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
XX atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
XX glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
XX osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
XX bacterial, fungal, parasitic, protozoal and helminthic infections and
XX cell proliferative disorders such as actinic keratosis, arteriosclerosis
XX and cancer including breast, bladder, bone marrow, brain and uterus
XX cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.
XX SQ Sequence 207 AA;

Query Match 22.8%; Score 124.5; DB 22; Length 207;
Best Local Similarity 36.0%; Pred. No. 3.1e-07;
Matches 36; Conservative 21; Mismatches 34; Indels 9; Gaps 4;
QY 13 ATNSNTATKNG----RYIYLEAFLEGAPWFTLKGLEHGEPLIISKVEEGKADTLSS 68
DB 77 atvaataseghahprvvelpktdeg---lgfninggkeqnsplivrsivpggvadr-hg 132
QY 69 KLAGDEVHINEVTLSSSRKE-AVSLVKGSYKTLRLVVR 107

DB 133 gikrgdqlsvngvsvegeqhekaavellikaagsvklvvr 172
RESULT 6
AAW70326
ID AAW70326 standard; Protein; 406 AA.
XX AC AAW70326;
XX 21-DEC-1998 (first entry)
XX Secreted protein DU123_1.
XX Secreted protein; DU123_1; human.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Misc-difference 390
FT /note= "encoded by AAY"
XX WO9838209-A2;
XX 03-SEP-1998.
XX 25-FEB-1998; 98WO-US03697.
XX 24-FEB-1998; 98US-0028724.
XX 26-FEB-1997; 97US-0805819.
XX (GEM) GENETICS INST INC.
XX Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D;
XX Racine LA, Spaulding V, Treacy M;
XX WPI: 1998-481139/41.
XX N-PSDB; AAV33198.
XX New isolated polynucleotide(s) and encoded polypeptide(s) -
XX obtained from human foetal kidney, adult colon, adult brain, foetal
XX brain and placenta cDNA libraries.
XX Claim 33; Page 80-81; 103pp; English.
XX This is the amino acid sequence of novel human secreted protein
XX DU123_1, as deduced from a full-length cDNA clone (see AAV33198)
XX obtained from a human foetal brain cDNA library. Database
XX searching revealed some similarity between DU123_1 and some known
XX sequences. The invention provides new isolated polynucleotides
XX (see AAV33198-99), from human foetal kidney, adult colon, adult brain,
XX foetal brain and placenta cDNA libraries, that code for secreted
XX proteins (see AAV70319-27). The clones can be used for recombinant
XX production of the polypeptides, which may have activities such as
XX e.g. nutritional activity, immunostimulant or immunosuppressive,
XX differentiation activity, cytokine and cell proliferation or
XX haematopoiesis regulating activity, tissue growth activity, activin
XX or inhibin activity, chemotactic or chemokinetic activity,
XX haemostatic and thrombolytic activity, receptor/ligand activity,
XX antiinflammatory activity, cadherin/tumour invasion suppressor
XX activity, tumour inhibition activity, or other activities.
XX SQ Sequence 406 AA;

Query Match 22.5%; Score 122.5; DB 19; Length 406;
Best Local Similarity 36.4%; Pred. No. 1.4e-06;
Matches 28; Conservative 16; Mismatches 30; Indels 3; Gaps 2;
QY 32 LEGAPGWFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVHIN-EVTLSSSRKE 90
DB 11 mtggapwgrlqggkqkqplqvakirndskas--ggglcegdvsvsngnncadltype 68

	Query Match	22.5%; Score 122.5; DB 20;	Length 913;
	Best Local Similarity	27.7%;	Pred. No. 4.5e-06;
	Matches 33;	Conservative 23;	Mismatches 44;
		Indels 19;	Gaps 8;


```

QY      6 EDFHKPSATLNSNTATKGRV-----IYLEAFLEGGAPWGFTLKGLGHEGEPLI 53
       :|||:: |:: ||:: :::: :|| |||:: ||:
Db    479 ddfhr--vtkggstedasqycdkndngdsylvlritpdedgkfglnlgvvdqkmplv 536
                               :|||:: |||:: |||:: |||:: |||:: |||::
QY     54 ISKYEEGGKADTLSSKLQAGDEVVHINEVTLS:-SRKEAVSLVKGSYKT----LRLVVR 107
       :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     537 vsrinpespadtcipkinegdqvlingrdisehtdhgvnmfikasreshrelalvir 595

```


RESULT 9

AAY07095

ID AAY07095 standard; Protein; 403 AA.

XX

AC AAY07095;

XX

DT 02-JUL-1999 (first entry)

XX

DE Colon cancer associated antigen precursor sequence.

XX

KW Cancer associated antigen; diagnosis; research; treatment; human;

KK breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

KW prostate cancer.

XX

OS Homo sapiens.

XX

PN WO9904265-A2.

[illegible]


```
OS Homo sapiens.
XX WO9904265-A2.
XX PN
XX PD
XX PF
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
XX PI Tureci O;
XX DR WPI; 1999-132448/11.
XX XX
XX PT New isolated cancer associated nucleic acids and polypeptides -
XX PT isolated using sera from cancer patients, used to develop products
XX PT for the diagnosis, monitoring or treatment of cancers
XX PS Disclosure; Page 656-658; 787pp; English.
XX CC The invention relates to a method for diagnosing a disorder characterised
XX CC by expression of a human cancer associated antigen precursor coded for by
XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX CC biological sample isolated from a subject with an agent that specifically
XX CC binds to the NAM, an expression product or a fragment of an expression
XX CC product complexed with an HLA molecule; and (b) determining the
XX CC interaction between the agent and the NAM or the expression product as a
XX CC determination of the disorder. The products and methods can be used in
XX CC the diagnosis, monitoring, research, or treatment of conditions
XX CC characterised by the expression of various cancer associated antigens.
XX CC The invention provides nucleic acid sequences and encoded polypeptides
XX CC which are cancer associated antigen precursors expressed in human breast
XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX CC lung cancer.
XX SQ Sequence 652 AA;

Query Match 22.4%; Score 122; DB 20; Length 652;
Best Local Similarity 41.4%; Pred. No. 3.2e-06;
Matches 29; Conservative 20; Mismatches 17; Indels 4; Gaps 3;

QY 39 GFTLKGGLEHGEPLIISKVEEGKADTLSSKLOAGDEVHINEVTLS-SRKEAVSLVK 97
| : : : | | | | | : : : | | : : : | | : : : | | : : : | : : :
Db 99 glsvrgglefcgclfishllkqgqadsvg--lqvqdeivringysissctheevinlir- 155
QY 98 SYKTLRLVWR 107
: | : : | |
Db 156 tkktvsikvr 165

RESULT 13
ID AAY07050
XX AAY07050 standard; Protein; 652 AA.
XX AC AAY07050;
XX XX
XX DT 02-JUL-1999 (first entry)
XX DE Renal cancer associated antigen precursor sequence.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX KW prostate cancer.

XX OS Homo sapiens.
XX PN WO9904265-A2.
XX PD
XX PF
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
XX PI Tureci O;
XX DR WPI; 1999-132448/11.
XX XX
XX PT New isolated cancer associated nucleic acids and polypeptides -
XX PT isolated using sera from cancer patients, used to develop products
XX PT for the diagnosis, monitoring or treatment of cancers
XX PS Disclosure; Page 656-658; 787pp; English.
XX CC The invention relates to a method for diagnosing a disorder characterised
XX CC by expression of a human cancer associated antigen precursor coded for by
XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX CC biological sample isolated from a subject with an agent that specifically
XX CC binds to the NAM, an expression product or a fragment of an expression
XX CC product complexed with an HLA molecule; and (b) determining the
XX CC interaction between the agent and the NAM or the expression product as a
XX CC determination of the disorder. The products and methods can be used in
XX CC the diagnosis, monitoring, research, or treatment of conditions
XX CC characterised by the expression of various cancer associated antigens.
XX CC The invention provides nucleic acid sequences and encoded polypeptides
XX CC which are cancer associated antigen precursors expressed in human breast
XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX CC lung cancer.
XX SQ Sequence 652 AA;

Query Match 22.4%; Score 122; DB 20; Length 652;
Best Local Similarity 41.4%; Pred. No. 3.2e-06;
Matches 29; Conservative 20; Mismatches 17; Indels 4; Gaps 3;

QY 39 GFTLKGGLEHGEPLIISKVEEGKADTLSSKLOAGDEVHINEVTLS-SRKEAVSLVK 97
| : : : | | | | | : : : | | : : : | | : : : | | : : : | : : :
Db 99 glsvrgglefcgclfishllkqgqadsvg--lqvqdeivringysissctheevinlir- 155
QY 98 SYKTLRLVWR 107
: | : : | |
Db 156 tkktvsikvr 165

RESULT 14
ID AAB55840
XX AAB55840 standard; Peptide; 86 AA.
XX AC AAB55840;
XX XX
XX DT 07-MAR-2001 (first entry)
XX DE PDZ encoded domain #10.
XX KW Hematopoietic cell; PDZ; PL; autoimmune disease; inflammation;
XX KW allergy; asthma; multiple sclerosis; cancer; infection.
```

XX	Synthetic.
OS	
XX	WO200069896-A2.
PN	
XX	23-NOV-2000.
PD	
XX	
PF	12-MAY-2000; 2000WO-US13161.
XX	
PR	14-MAY-1999; 99US-0134114.
XX	14-MAY-1999; 99US-0134117.
PR	14-MAY-1999; 99US-0134118.
PR	21-OCT-1999; 99US-0160860.
PR	29-OCT-1999; 99US-0162498.
PR	13-DEC-1999; 99US-0170453.
PR	14-JAN-2000; 2000US-0176195.
PR	14-FEB-2000; 2000US-0182296.
PR	11-APR-2000; 2000US-0186460.
PR	11-APR-2000; 2000US-0196527.
XX	
PA	(ARBO-) ARBOR VITA CORP.
XX	
PI	Lu PS:
DR	WPI: 2001-080245/09.
XX	
PT	Modulating a biological function of an endothelial cell or
PT	hematopoietic cell, useful for treating autoimmune diseases and
PT	infectious diseases, by administering an antagonist that inhibits
PT	binding between a PDZ protein and a PL protein -
PS	Disclosure; Page 28-43; 141pp; English.
XX	
CC	The present invention relates to a new method for modulating a
CC	biological function of an endothelial cell or hematopoietic cell. The
CC	method involves introducing into a cell, an antagonist that inhibits
CC	binding between a PDZ protein and a PL protein. The inhibitor is used
CC	to treat a disease mediated by hematopoietic cells, e.g. autoimmune
CC	disease. It may also be used to prevent transplantation rejection of
CC	a solid organ transplant. The method may also be used in the treatment
CC	of inflammation, allergy, inflammatory bowel diseases, ulcerative
CC	colitis, ileitis, psoriasis, asthma, atopic dermatitis, autoimmune
CC	diseases (e.g. rheumatoid arthritis, multiple sclerosis,
CC	insulin-dependent diabetes, Hashimoto thyroiditis, osteoarthritis,
CC	graft rejection, transplantation rejection), atherosclerosis, cancers,
CC	infectious diseases, ischemia, vasculitis and Crohn's disease.
XX	
SQ	Sequence 86 AA;
XX	
Query Match	21.6%; Score 117.5; DB 22; Length 86;
Best Local Similarity	35.1%; Pred. No. 6.6e-07;
Matches 27; Conservative 18; Mismatches 29; Indels 3; Gaps 2;	
QY	32 LEGAAMGFTIKGLSEHEPLELISKEEGGRADTLSSKLOAGDEVHINVEYLSS-SRKE 90 : : : : : : D 8 lvgpepwgfirigqgdfmpltlsiskdgysk--eqanvrigdvslsdgnaagmthle 65 : : : : : : : : : : : : : :
OY	91 AVSLVKGSYKTILRLVVR 107 : : : : : : : : : : : : : :
Db	66 agnkikgctgslnmtlg 82
XX	
RESULT 15	
ID	AAB57631 standard; Protein: 86 AA.
XX	
AC	AAB57631;
XX	
DT	12-MAR-2001 (first entry)
XX	
DE	Human LIM protein PDZ domain.
XX	

```

OS      Homo sapiens.
XX
XX      MOZ00069897-A2.
XX
XX      23-NOV-2000.
XX
XX      12-MAY-2000; 2000MO-U0S13166.
XX
XX      14-MAY-1999; 99US-0134114.
XX      14-MAY-1999; 99US-0134117.
XX      14-MAY-1999; 99US-0134118.
XX      21-OCT-1999; 99US-0160860.
XX      29-OCT-1999; 99US-0162498.
XX      13-DEC-1999; 99US-0170453.
XX      14-JAN-2000; 2000US-0176195.
XX      14-FEB-2000; 2000US-0182296.
XX      11-APR-2000; 2000US-0196460.
XX      11-APR-2000; 2000US-0196527.
XX
XX      (ARBO-) ARBOR VITA CORP.
XX
XX      Lu PS:
XX
XX      WPI: 2001-025003/03.
XX
PT      New inhibitors of binding of a PDZ protein and PL protein for
PT      inhibiting T cell-mediated response by hematopoietic cells, or for
PT      treating diseases characterized by inflammatory and humoral immune
PT      responses, e.g. inflammation, cancer
XX
XX      Disclosure; Page 32; 139pp; English.
XX
CC      The present invention relates to a method for modulating a biological
CC      function of an endothelial cell or hematopoietic cell, comprises
CC      introducing into a cell an antagonist that inhibits binding between a
CC      PDZ domain protein and a PL domain protein to result in inhibition of
CC      leukocyte activation. The present sequence is a PDZ domain. PDZ domains
CC      of Drosophila are named after three prototypical proteins: PSD95,
CC      Drosophila large disc protein and Zonula Occludin I protein. PDZ domain
CC      proteins are involved in synapse formation by organising transmembrane
CC      neurotransmitter receptors through intracellular interactions. The
CC      inhibitors identified by the present invention can be used to treat a
CC      disease mediated by hematopoietic cells, e.g. autoimmune disease,
CC      inflammation, allergy (e.g. drug allergies), inflammatory bowel diseases,
CC      ulcerative colitis, ileitis, psoriasis, respiratory allergic diseases
CC      (e.g. asthma), atopic dermatitis, autoimmune diseases (e.g. rheumatoid
CC      arthritis, multiple sclerosis, insulin-dependent diabetes, Hashimoto
CC      thyroiditis, osteoarthritis), atherosclerosis, cancers, infectious
CC      diseases (e.g. viral infection), ischaemia, vasculitis and Crohn's
CC      disease. The inhibitors can also be used to prevent transplantation
CC      rejection of a solid organ transplant.
XX
XX      Sequence      86 AA:
SQ

```

Db 66 agnkikgetgslnmtlq 82

Search completed: January 31, 2002, 20:56:12
Job time: 80 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:57:10 ; Search time 29.66 Seconds
(without alignments)
81.941 Million cell updates/sec

Title: US-09-641-831-4

Perfect score: 545
Sequence: 1 MMRTEDEFRKPSATLNSNTA.....KEAVSLVKGSKYTLRLVRS 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	21.5	75	US-09-100-804-21	Sequence 21, Appl
2	113.5	20.8	327	US-08-739-485-10	Sequence 10, Appl
3	112.5	20.6	233	US-09-151-611-1	Sequence 1, Appl
4	112.5	20.6	233	US-09-370-102-1	Sequence 1, Appl
5	112.5	20.6	329	US-08-739-485-3	Sequence 3, Appl
6	112.5	20.6	330	US-08-739-485-11	Sequence 11, Appl
7	109.5	20.1	316	US-08-739-485-5	Sequence 5, Appl
8	109	20.0	297	US-09-151-611-3	Sequence 3, Appl
9	109	20.0	297	US-09-370-102-3	Sequence 3, Appl
10	106.5	19.5	223	US-09-124-238A-34	Sequence 34, Appl
11	106.5	19.5	457	US-09-124-238A-1	Sequence 1, Appl
12	106.5	19.5	457	US-09-124-238A-10	Sequence 10, Appl
13	104.5	19.2	454	US-08-166-316-2	Sequence 2, Appl
14	103.5	19.0	73	US-09-100-804-28	Sequence 28, Appl
15	95	17.4	182	US-09-045-632-29	Sequence 29, Appl
16	95	17.4	283	US-09-045-632-27	Sequence 27, Appl
17	95	17.4	498	US-09-045-632-30	Sequence 30, Appl
18	95	17.4	502	US-09-045-632-24	Sequence 24, Appl
19	95	17.4	505	US-08-123-161A-14	Sequence 14, Appl
20	95	17.4	505	US-08-483-278-14	Sequence 14, Appl
21	95	17.4	541	US-09-045-632-36	Sequence 36, Appl
22	95	17.4	599	US-09-045-632-28	Sequence 28, Appl
23	95	17.4	602	US-09-045-632-20	Sequence 20, Appl
24	95	17.4	642	US-09-045-632-35	Sequence 35, Appl
25	95	17.4	702	US-09-045-632-15	Sequence 15, Appl
26	95	17.4	818	US-09-045-632-25	Sequence 25, Appl
27	95	17.4	861	US-09-045-632-34	Sequence 34, Appl

28	95	17.4	918	3	US-09-045-632-21	Sequence 21, Appl
29	95	17.4	961	3	US-09-045-632-33	Sequence 33, Appl
30	95	17.4	1018	3	US-09-045-632-16	Sequence 16, Appl
31	95	17.4	1061	3	US-09-045-632-32	Sequence 32, Appl
32	95	17.4	1112	3	US-09-045-632-2	Sequence 2, Appl
33	95	17.4	1112	3	US-09-045-632-3	Sequence 3, Appl
34	93.5	17.2	80	3	US-08-545-860D-54	Sequence 54, Appl
35	93.5	17.2	80	5	PCT-US94-04496-54	Sequence 54, Appl
36	92.5	17.0	73	3	US-09-100-804-25	Sequence 25, Appl
37	92.5	17.0	74	3	US-09-100-804-34	Sequence 34, Appl
38	88	16.1	82	3	US-09-045-632-9	Sequence 9, Appl
39	88	16.1	398	3	US-09-045-632-31	Sequence 31, Appl
40	88	16.1	441	3	US-09-045-632-37	Sequence 37, Appl
41	87.5	16.1	75	3	US-09-100-804-22	Sequence 22, Appl
42	87.5	16.1	79	3	US-09-100-804-27	Sequence 27, Appl
43	87.5	16.1	86	3	US-08-545-860D-53	Sequence 53, Appl
44	87.5	16.1	86	5	PCT-US94-04496-53	Sequence 53, Appl
45	84	15.4	2037	4	US-09-306-998-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-100-804-21
Sequence 21, Application US/09100804
Patent No. 6066472
GENERAL INFORMATION:
APPLICANT: GONZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESSON-WELSH, LENA
APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NOCOTOTIDE SEQUENCES FOR NOVEL PROTEIN
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,804
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/596,291
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-100-804-21

Query Match 21.5%; Score 117; DB 3; Length 75;
Best Local Similarity 38.8%; Pred. No. 1.3e-07;
Matches 26; Conservative 12; Mismatches 27; Indels 2; Gaps 2;

QY 33 EGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLOAGDEVVHINEVTLSS-SRKEA 91
Db 2 EDGRP-GFNKGGVQKNPLVSRINPSSPADICPKLNEGDIIVLINGRDISEHTDQV 60

QY 92 VSLVKGS 98
Db 61 VMFIKAS 67

RESULT 2
US-08-739-485-10
Sequence 10, Application US/08739485
Patent No. 5863898

GENERAL INFORMATION:
APPLICANT: Goll, Surya K.
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/739,485
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0142 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1020151
US-08-739-485-10

Query Match 20.8%; Score 113.5; DB 2; Length 327;
Best Local Similarity 39.5%; Pred. No. 3.1e-06;
Matches 30; Conservative 11; Mismatches 32; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLOAGDEVVHIN-EVTLSSRKE 90
Db 8 LOGPGWGERLVGGKDFEQPLAISRTVPGSKA--ATANLCIGDLITADIGEDTSSNTHLE 65
QY 91 AVSLVKGSYKTLRLVV 106
Db 66 AQNKINGCYDNTMTLV 81

RESULT 3
US-09-151-611-1
Sequence 1, Application US/09151611
Patent No. 5958731
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Au-Young, Janice
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
FILE REFERENCE: PF-0599 US
CURRENT APPLICATION NUMBER: US/09/151,611
CURRENT FILING DATE: 1998-09-11
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 233
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 1974337
US-09-151-611-1

Query Match 20.6%; Score 112.5; DB 2; Length 233;
Best Local Similarity 33.0%; Pred. No. 2.5e-06;
Matches 33; Conservative 22; Mismatches 36; Indels 9; Gaps 4;

QY 13 ATLSNTATKG----RYIYLEAFLEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSS 68
Db 92 ATVAFAASEGSHPRVVELPKTDEG---LGFNVMGKEQNSPIYIRIIPGGVAER-HG 147

QY 69 KLOAGDEVVHINEVTLSSRKE-AVSLVKGSYKTLRLVVR 107
Db 148 GLKRGDQLLSVNGSVSVEGEHHEKRAVELLKAADSVKLVR 187

RESULT 4
US-09-370-102-1
Sequence 1, Application US/09370102
Patent No. 6265547
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Au-Young, Janice
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
FILE REFERENCE: PF-0599 US
CURRENT APPLICATION NUMBER: US/09/370,102
CURRENT FILING DATE: 1999-08-06
EARLIER APPLICATION NUMBER: 09/151,611
EARLIER FILING DATE: 1998-09-11
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 233
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 1974337
US-09-370-102-1

Query Match 20.6%; Score 112.5; DB 4; Length 233;
Best Local Similarity 33.0%; Pred. No. 2.5e-06;
Matches 33; Conservative 22; Mismatches 36; Indels 9; Gaps 4;

QY 13 ATINSNTATK-----RYIXLEAFLEGAPMGFTLKGLEHGEPLITSKEEGKADTLSS 68
 DB 92 ATVAFAAEGSHSHPRVELLPRTDEG---LGFVVMGKQNSPIYISRLIPGVABR-HQ 147
 QY 69 KQAGDEVHINEVYLSSRKE-AVSLVKGSKYKTLRLV 107
 DB 148 GLKRGDOLLVNGSVGEHHEKAVELLAAKDSVKLV 187

RESULT 5

US-08-739-485-3
 ; Sequence 3, Application US/08739485
 ; Patent No. 5863898
 ; GENERAL INFORMATION:
 ; APPLICANT: Goll, Surya K.
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Bandman, Olga
 ; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: US
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/739,485
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0142 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 329 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Consensus
 ; CLONE: Consensus
 ; US-08-739-485-3

Query Match 20.6%; Score 112.5; DB 2; Length 329;
 Best Local Similarity 39.5%; Pred. No. 4.2e-06;
 Matches 30; Conservative 11; Mismatches 32; Indels 3; Gaps 2;

QY 32 LEGAPMGFTLKGLEHGEPLITSKEEGKADTLSSKLAGDEVHIN-EVTLSSSRKE 90
 DB 8 LQGGPMPGFRVLVGGKDEQPLAISRVPGSKA--ALANLCIGDVTALDSENTSMTHLE 65
 QY 91 AVSLVKGSKYKTLRLV 106
 DB 66 AONRIKCGDHLTLV 81

RESULT 6

US-08-739-485-11
 ; Sequence 11, Application US/08739485
 ; Patent No. 5863898
 ; GENERAL INFORMATION:
 ; APPLICANT: Goll, Surya K.
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Bandman, Olga
 ; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: US
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/739,485
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0142 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 330 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; ORIGINAL SOURCE:
 ; HAPLOTYPE: GenBank
 ; IMMEDIATE SOURCE:
 ; LIBRARY: 887580
 ; US-08-739-485-11

Query Match 20.6%; Score 112.5; DB 2; Length 330;
 Best Local Similarity 40.8%; Pred. No. 4.2e-06;
 Matches 31; Conservative 9; Mismatches 33; Indels 3; Gaps 2;

QY 32 LEGAPMGFTLKGLEHGEPLITSKEEGKADTLSSKLAGDEVHIN-EVTLSSSRKE 90
 DB 7 LRGPSPMGFRVLVGGKDEQPLAISRVHNSKA--ALALCGDSIQALNGSTELMTHLE 64
 QY 91 AVSLVKGSKYKTLRLV 106
 DB 65 AONRIKCGDHLTLV 80

RESULT 7
 US-08-739-485-5
 ; Sequence 5, Application US/08739485
 ; Patent No. 5863898
 ; GENERAL INFORMATION:
 ; APPLICANT: Goll, Surya K.
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Bandman, Olga
 ; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:

US-08-739-485-5
 ; Sequence 5, Application US/08739485
 ; Patent No. 5863898
 ; GENERAL INFORMATION:
 ; APPLICANT: Goll, Surya K.
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Bandman, Olga
 ; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/739,485

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0142 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 316 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: Consensus

CLONE: Consensus

US-08-739-485-5

Query Match

Best Local Similarity 20.1%; Score 109.5; DB 2; Length 316;

Matches 26; Conservative 17; Mismatches 30; Indels 3; Gaps 2;

QY 32 LEGGAPWFTLGGLEHGEPLIISKVEEGKADTLSSKLQAGDEVVHNEV-TLSSSRKE 90

DB 7 LCPAPWGRFLSGIDFNQPLVITRITPGSKA--AAANLCPGDVILADFGFTESMTHAD 64

QY 91 AVSLVRGSKYKTLRLVV 106

DB 65 AQDRIKAAAHQLCLKI 80

RESULT 8

US-09-151-611-3

Sequence 3, Application US/09151611

Patent No. 5958731

GENERAL INFORMATION:

APPLICANT: Yue, Henry

APPLICANT: Au-Young, Janice

APPLICANT: Patterson, Chandra

TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN

FILE REFERENCE: PF-0599 US

CURRENT APPLICATION NUMBER: US/09/151,611

CURRENT FILING DATE: 1998-09-11

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PERL Program

SEQ ID NO 3

LENGTH: 297

TYPE: PRT

ORGANISM: Caenorhabditis elegans

FEATURE: -

OTHER INFORMATION: g1685067

US-09-151-611-3

Query Match

Best Local Similarity 20.0%; Score 109; DB 2; Length 297;

Matches 30; Conservative 23; Mismatches 38; Indels 8; Gaps 3;

QY 13 ATLNSTATKG---RYIYLEAFLEGGAPWFTLGGLEHGEPLIISKVEEGKADTLSS 68

DB 184 ATVAAPAAAEHGAHPRIVELPKTDG---LGFNVMGKQNSPIYISRIIPGGVADR-HG 239

QY 69 KIQAGDEVVHNEVTLSRSSRKEAVSLVKGSYKTLRLVVR 107

DB 240 GLKRGDQLIAVNGVNEAECHERKAVDLLKSAVGSVKLVIR 278

RESULT 9

US-09-370-102-3

Sequence 3, Application US/09370102

Patent No. 6265547

GENERAL INFORMATION:

APPLICANT: Yue, Henry

APPLICANT: Au-Young, Janice

APPLICANT: Patterson, Chandra

TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN

FILE REFERENCE: PF-0599 US

CURRENT APPLICATION NUMBER: US/09/370,102

CURRENT FILING DATE: 1999-08-06

EARLIER APPLICATION NUMBER: 09/151,611

EARLIER FILING DATE: 1998-09-11

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PERL Program

SEQ ID NO 3

LENGTH: 297

TYPE: PRT

ORGANISM: Caenorhabditis elegans

FEATURE: -

OTHER INFORMATION: g1685067

US-09-370-102-3

Query Match

Best Local Similarity 20.0%; Score 109; DB 4; Length 297;

Matches 30; Conservative 23; Mismatches 38; Indels 8; Gaps 3;

QY 13 ATLNSTATKG---RYIYLEAFLEGGAPWFTLGGLEHGEPLIISKVEEGKADTLSS 68

DB 184 ATVAAPAAAEHGAHPRIVELPKTDG---LGFNVMGKQNSPIYISRIIPGGVADR-HG 239

QY 69 KIQAGDEVVHNEVTLSRSSRKEAVSLVKGSYKTLRLVVR 107

DB 240 GLKRGDQLIAVNGVNEAECHERKAVDLLKSAVGSVKLVIR 278

RESULT 10

US-09-124-238A-34

Sequence 34, Application US/09124238A

Patent No. 6300127

GENERAL INFORMATION:

APPLICANT: Hair, Gregory A.

APPLICANT: Boden, Scott D.

TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,

FILE REFERENCE: Expression Systems

FILE REFERENCE: 06148.0115

CURRENT APPLICATION NUMBER: US/09/124,238A

CURRENT FILING DATE: 1998-07-29

PRIOR APPLICATION NUMBER: 60/054,219

PRIOR FILING DATE: 1997-07-30

PRIOR APPLICATION NUMBER: 60/080,407

PRIOR FILING DATE: 1998-04-02

NUMBER OF SEQ ID NOS: 36

SOFTWARE: MS Word

SEQ ID NO 34

LENGTH: 223

TYPE: PRT

Fri Feb 1 08:56:30 2002

us-09-641-831-4.ra1

A: Note: PSD-95/SAP90-related gene 2 (chapsyn isoform2)
C: Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase homology
F: 198-276/Domain: GLGF domain homology <GLG>
F: 543-601/Domain: SH3 homology <SH3>
F: 663-840/Domain: guanylate kinase homology <GKI>

Query Match 23.1%; Score 126; DB 2; Length 852;
Best Local Similarity 32.7%; Pred. No. 5.3e-05;
Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;

Qy 10 KPSATLNSNTATKG--RYIYLEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLS 67
Db 404 QPSVTQRAISLEGEPRKVVHLH---KGSTGLGFNIVGG-EDGEGIFVIFLAGGPAD-LS 458

Qy 68 SKLQAGDEVVHINEVTL-SSSRKEAVSLVKGSKYKTLRLVVR 107
Db 459 GELQRGDQILSVNGIDLRGASHEQAALAKGAGQTVITIAQ 499

RESULT 3
G01974
channel associated protein of synapse - human
C: Species: Homo sapiens (man)
C: Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 21-Jan-2000
C: Accession: G01974
R: Kim, E.; Cho, K.
submitted to the EMBL Data Library, July 1995
A: Reference number: G08966
A: Accession: G01974
A: Status: preliminary; translated from GB/EMBL/DBDJ
A: Molecule type: mRNA
A: Residues: 1-870 <KIM>
A: Cross-references: EMBL:U32376; NID: g1463025; PID: g1036790
C: Genetics:

Query Match 23.1%; Score 126; DB 2; Length 870;
Best Local Similarity 32.7%; Pred. No. 5.5e-05;
Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;

Qy 10 KPSATLNSNTATKG--RYIYLEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLS 67
Db 404 QPSMTLQRAVSLGEPRKVVHLH---KGSTGLGFNIVGG-EDGEGIFVIFLAGGPAD-LS 458

Qy 68 SKLQAGDEVVHINEVTL-SSSRKEAVSLVKGSKYKTLRLVVR 107
Db 459 GELQRGDQILSVNGIDLRGASHEQAALAKGAGQTVITIAQ 499

RESULT 4
A41109
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 3 - human
N: Alternate names: PTPH1
C: Species: Homo sapiens (man)
C: Date: 27-Mar-1992 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C: Accession: A41109; I55698
R: Yang, Q.; Tonks, N.K.
Proc. Natl. Acad. Sci. U.S.A. 88, 5949-5953, 1991
A: Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with homology to the SH-PTPase
A: Reference number: A41109; MUID: 91296738
A: Accession: A41109
A: Molecule type: mRNA
A: Residues: 1-913 <YAN>
A: Cross-references: GB: M64572; NID: g179912; PIDN: AAA35647.1; PID: g179913
R: Ikuta, S.; Itoh, F.; Hinoda, Y.; Toyota, M.; Makiguchi, Y.; Imai, K.; Yachi, A.
J. Gastroenterol. 29, 727-732, 1994
A: Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA

A: Reference number: I55698; MUID: 95179278
A: Accession: I55698
A: Status: translated from GB/EMBL/DBDJ
A: Molecule type: mRNA
A: Residues: 899-913 <RES>
A: Cross-references: GB: S76309; NID: g913165; PIDN: AAB33583.1; PID: g913166
C: Genetics:

A: Gene: GDB: PTPN3
A: Cross-references: GDB: I131386; OMIM: 176877
A: Map position: 9q31-q31
C: Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology
C: Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase; protein-4.1 membrane-binding domain homology <B41>
F: 31-308/Domain: GLGF domain homology <GLG>
F: 516-590/Domain: GLGF domain homology <GLG>
F: 670-890/Domain: protein-tyrosine-phosphatase homology <PTP>
F: 842/Active site: Cys (phosphocysteine intermediate) #status predicted
F: 848/Binding site: substrate phosphate (Arg) #status predicted

Query Match 22.5%; Score 122.5; DB 1; Length 913;
Best Local Similarity 27.7%; Pred. No. 0.00013;
Matches 33; Conservative 23; Mismatches 44; Indels 19; Gaps 4;

Qy 6 EDPHKPSATLNSNTATKGRY-----IYLEAFLEGAPWGFTLKGLEHGEPLI 53
Db 479 DDFHR--VTGGSTEDASQYCDKNDGDSVLVIRITPDGKFGFNKGGVDKMKPLV 536

Qy 54 ISKVEEGKADTSSKLOAGDEVVHINEVTLSS-SRKEAVSLVKGSKYK-----LRLVVR 107
Db 537 VSRINPESPADTCIPKLNQEGQIVLNGRDISETHDQVVMFIKASRSHSRELALVIR 595

RESULT 5
T42372
probable guanylate kinase (EC 2.7.4.8) 1, membrane-associated, splice form b - mouse
C: Species: Mus musculus (house mouse)
C: Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C: Accession: T42372
R: Dobrosotskaya, I.; Guy, R.K.; James, G.L.
J. Biol. Chem. 272, 31589-31597, 1997
A: Title: MAGI-1: a membrane-associated guanylate kinase with a unique arrangement of intracellular domain
A: Reference number: 222139; MUID: 98058950
A: Accession: T42372
A: Status: preliminary; translated from GB/EMBL/DBDJ
A: Molecule type: mRNA
A: Residues: 1-1171 <DOB>
A: Cross-references: EMBL: AF027503; NID: g2702346; PID: g2702347; PIDN: AAB91995.1
A: Experimental source: strain C57 Black/6 x CBA
C: Genetics:

A: Gene: Magi-1
C: Superfamily: WW repeat homology
C: Keywords: alternative splicing; phosphotransferase
F: 300-337/Domain: WW repeat homology <WWRL>
F: 347-384/Domain: WW repeat homology <WWRL2>

Query Match 22.0%; Score 120; DB 2; Length 1171;
Best Local Similarity 28.0%; Pred. No. 0.00033;
Matches 33; Conservative 20; Mismatches 45; Indels 20; Gaps 4;

Qy 9 HKPS--ATLNSNTATKGRY-----IYLEAFLEGAPWGFTLKGLEHGE 50
Db 998 HAPSOQGTQETRTTKPKQDSQFQAEQDFYVVELERGAKGFGSLRGGRYNM 1057

Qy 51 PLIISKVEEGKADTSSKLOAGDEVVHINEVTLSSSRKEAVSLVKGSKYKTLRLVVR 107
Db 1058 DLYVLRLEADGPAER-CGKMRIGDSEILENGETTKMKHSRAIELIKNGRRVRLFLR 1114

RESULT 6
JB0209
brain-specific angiogenesis inhibitor-associated protein 1 - human
N: Alternate names: BA11-associated protein 1; BAP1 [misnomer]

RESULT 10
I38757
homolog of Drosophila discs large protein, isoform 1 - human
C:Species: Homo sapiens (man)
C>Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: I38757
R:Luc, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila disc large protein
A:Reference number: I38756; MUID:95024052
A:Accession: I38757
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-904 <RES>
A:Cross-references: EMBL:U13897; NID:q558437; PIDN:AAA50599.1; PID:q558438
C:Genetics:
A:Gene: GDB:DLG1
A:Cross-references: GDB:393278; OMIM:601014
A:Map position: 3q29-3q29
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase homology
C:Keywords: alternative splicing; duplication
F:229-307/Domain: GLGF domain homology <GLG1>
F:324-402/Domain: GLGF domain homology <GLG2>
F:588-646/Domain: SH3 homology <SH3>
F:715-892/Domain: guanylate kinase homology <GKI>

Query Match 20.4%; Score 111; DB 2; Length 904;
Best Local Similarity 30.3%; Pred. No. 0.0021;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

QY 7 DPH-KPSATLNSNTATKGRYYL-EAFL-----EGGAPWGFTLKGGLHG 49
DB 428 DNHVSPSSFLGQTSPASPARYSPVSKAVLGDDDEITREPRKVVHLRGSTGLGFNIYVG-EDG 486
QY 50 EPLIISKVEEGKADTLSSKLGQAGDEVVHINEVTL-SSSRKEAVSLVKGSKYKTLRLVVR 107
DB 487 EGIFISFILAGGPAD-LSGELKRGDRIISVNSVDLRAASHEQAAALKNAGQAVTIVAQ 544

RESULT 11
I38756
homolog of Drosophila discs large protein, isoform 2 - human
C:Species: Homo sapiens (man)
C>Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: I38756
R:Luc, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila disc large protein
A:Reference number: I38756; MUID:95024052
A:Accession: I38756
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-926 <RES>
A:Cross-references: EMBL:U13896; NID:q558435; PIDN:AAA50598.1; PID:q558436
C:Genetics:
A:Gene: GDB:DLG1
A:Cross-references: GDB:393278; OMIM:601014
A:Map position: 3q29-3q29
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase homology
C:Keywords: alternative splicing; duplication
F:229-307/Domain: GLGF domain homology <GLG1>
F:324-402/Domain: GLGF domain homology <GLG2>
F:588-646/Domain: SH3 homology <SH3>
F:737-914/Domain: guanylate kinase homology <GKI>

Query Match 20.4%; Score 111; DB 2; Length 926;
Best Local Similarity 30.3%; Pred. No. 0.0022;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

QY 7 DPH-KPSATLNSNTATKGRYYL-EAFL-----EGGAPWGFTLKGGLHG 49

Db 428 DNHVSPSSFLGQTSPASPARYSPVSKAVLGDDDEITREPRKVVHLRGSTGLGFNIYVG-EDG 486
QY 50 EPLIISKVEEGKADTLSSKLGQAGDEVVHINEVTL-SSSRKEAVSLVKGSKYKTLRLVVR 107
DB 487 EGIFISFILAGGPAD-LSGELKRGDRIISVNSVDLRAASHEQAAALKNAGQAVTIVAQ 544

RESULT 12
T27179
hypothetical protein Y54G11A.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27179
R:Wallis, J.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z20322
A:Accession: T27179
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-317 <WIL>
A:Cross-references: EMBL:AL034488; NID:el359895; PIDN:CAA22459.1; CBSP:Y54G11A.10
A:Experimental source: clone Y54G11A
C:Genetics:
A:Gene: CBSP:Y54G11A.10
A:Introns: 55/3; 108/3; 175/3; 228/2; 253/3

Query Match 20.3%; Score 110.5; DB 2; Length 317;
Best Local Similarity 31.0%; Pred. No. 0.00066;
Matches 31; Conservative 24; Mismatches 36; Indels 9; Gaps 4;

QY 13 ATLNSNTATKPGK---RYIYLEAFLEGAPWGFTLKGGLHGEPLIISKVEEGKADTLSS 68
DB 184 ATVAFAAAEGHAHPRIPELPTDQG---LGFNVMGKGKQNSPIYISRIIPGGVADR-HG 239
QY 69 KIQAGDEVVHINEVTLSSS-RKEAVSLVKGSKYKTLRLVVR 107
DB 240 GLKRGDQLIANGVNVVEAECEKAEKAVDLLASAVGSVKLVIR 279

RESULT 13
A55050
enigma - human
C:Species: Homo sapiens (man)
C>Date: 18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A55050
R:Wu, R.Y.; Gill, G.N.
J. Biol. Chem. 269, 25085-25090, 1994
A:Title: LIM domain recognition of a tyrosine-containing tight turn.
A:Reference number: A55050; MUID:95014287
A:Accession: A55050
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-455 <WUA>
A:Cross-references: GB:I35240; NID:q561636; PIDN:AAC37565.1; PID:q561637
C:Superfamily: LIM metal-binding repeat homology; GLGF domain homology
F:9-81/Domain: GLGF domain homology <GLG>
F:280-330/Domain: LIM metal-binding repeat homology <LIM1>
F:339-389/Domain: LIM metal-binding repeat homology <LIM2>
F:398-450/Domain: LIM metal-binding repeat homology <LIM3>

Query Match 19.5%; Score 106.5; DB 2; Length 455;
Best Local Similarity 44.4%; Pred. No. 0.0027;
Matches 28; Conservative 7; Mismatches 19; Indels 9; Gaps 2;

QY 32 LEGAPWGFTLKGGLHGEPLIISKVEEGKADTLSSKLGQAGDEVVHINEVTLSSSRKEA 91
DB 8 LEGPAPWGFTLKGGLHGEPLIISKVEEGKADTLSSSRTPGKA-----AQAG---VAVGDMVLSDGENA 58
QY 92 VSL 94
||

Db 59 GSL 61

Job time: 65 sec

RESULT 14

T47134

hypothetical protein DKFZp76112312.1 - human

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47134

R:Ansoyge, W.; Wilkner, U.; Mewes, H.W.; Well, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24375

A:Accession: T47134

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-517 <AAA>

A:Cross-references: EMBL:AL161971

A:Experimental source: adult amygdala; clone DKFZp76112312

C:Genetics:

A>Note: DKFZp76112312.1

Query Match 19.4%; Score 105.5; DB 2; Length 517;

Best Local Similarity 26.5%; Pred. No. 0.004; Mismatches 37; Indels 15; Gaps 3;

Matches 27; Conservative 23; Mismatches 37; Indels 15; Gaps 3;

OY 6 EDFHPSATLNSNTATKGRVYILEAFLEGGAPWFTLKGLEHGEPLIISKVEGGKADT 65

DB 49 EPFYSGERVTVTRQTVGFE-----GLSIKGAEHNPVYVSKISKEGRAE 94

OY 66 ISSKIQAGDEVVHINEVTLSSR-KEAVSLVKGSKTLELVV 106

DB 95 LSGLEFGDAILQINGINVRKCRHEVQVLRNAGEVTLTV 136

RESULT 15

A45436

synapse-associated protein SAP90 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000

C:Accession: A45436

R:Kistner, U.; Menzel, B.M.; Vep, R.W.; Cases-Langhoff, C.; Garner, A.M.; Appeltauer, U.

J. Biol. Chem. 268, 4580-4583, 1993

A:Title: SAP90, a rat presynaptic protein related to the product of the Drosophila tumor

A:Reference number: A45436; MUID:93186749

A:Accession: A45436

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-720 <KIS>

A:Experimental source: brain

A>Note: sequence extracted from NCBI backbone (NCBIP:126555)

C:Superfamily: discs-large tumor suppressor; GIGF domain homology; guanylate kinase hom

F:70-148/Domain: GIGF domain homology <GIGI>

F:165-243/Domain: GIGF domain homology <GIG2>

F:435-493/Domain: SH3 homology <SH3>

F:531-708/Domain: guanylate kinase homology <GKI>

Query Match 19.2%; Score 104.5; DB 2; Length 720;

Best Local Similarity 34.7%; Pred. No. 0.0076;

Matches 26; Conservative 20; Mismatches 26; Indels 3; Gaps 3;

OY 34 GGAPWFTLKGLEHGEPLIISKVEGGKADTSSKIQAGDEVVHINEVTL-SSSRKEAV 92

DB 319 GSTGLGFNIYVG-EDGRIPIFTLAGGPAD-LSGELRKGDQILSVNGVDLRNASHQAA 376

OY 93 SLVKGSKTLELVV 107

DB 377 IALKNAGQTVTIIAQ 391

Search completed: January 31, 2002, 20:59:17

us-09-641-831-4.rpr

Fri Feb 1 08:56:32 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:05:34 ; Search time 72.88 Seconds
(without alignments)
216.759 Million cell updates/sec

Title: US-09-641-831-4

Perfect score: 545
Sequence: 1 MRRTEDEHFKPSATLNSNTA.....KEAVSLYKSGYKTLRLVYRS 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP:REMBL_17:*
2: SP:archaea:*
3: SP:bacteria:*
4: SP:fungi:*
5: SP:human:*
6: SP:invertebrate:*
7: SP:mammal:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:protent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	409	75.0	1986	11	Q9QXN0
2	262.5	48.2	1502	4	Q9ULH8
3	134.5	24.7	1454	4	Q9H4G2
4	128	23.5	39	4	Q9H4T0
5	125.5	23.0	298	13	Q9PU46
6	124.5	22.8	207	4	Q9HAP6
7	123.5	22.7	207	11	Q9Z252
8	123.5	22.7	207	11	Q88951
9	123.5	22.7	234	11	Q9D201
10	123.5	22.7	315	13	Q9PU47
11	123	22.6	171	5	017458
12	122	22.4	403	4	Q9UPC3
13	122	22.4	548	11	Q9ES65
14	122	22.4	552	4	Q9UM04
15	122	22.4	652	4	Q9Y6N9
16	122	22.4	910	11	Q9ES64
17	121	22.2	493	5	Q9W443
18	120	22.0	521	4	Q9UM17
19	120	22.0	1171	11	Q54893

ALIGNMENTS

20	118.5	21.7	1256	4	Q75085	075085 homo sapien
21	117.5	21.6	596	4	Q60705	060705 homo sapien
22	116.5	21.4	400	5	Q9V7F2	Q9V7F2 drosophila
23	115.5	21.2	195	5	Q9VB77	Q9VB77 drosophila
24	115.5	21.2	197	4	Q9NUP9	Q9NUP9 homo sapien
25	115.5	21.2	197	11	Q88952	Q88952 mus musculus
26	114.5	21.0	337	11	Q9QYNO	Q9QYNO mus musculus
27	114.5	21.0	591	11	Q9QYNO2	Q9QYNO2 mus musculus
28	113.5	20.8	138	11	Q9D130	Q9D130 mus musculus
29	113.5	20.8	283	4	Q9Y4Z5	Q9Y4Z5 homo sapien
30	113.5	20.8	288	11	Q9R021	Q9R021 mus musculus
31	113.5	20.8	288	11	Q9WVH1	Q9WVH1 mus musculus
32	113.5	20.8	470	4	Q9Y4Z3	Q9Y4Z3 homo sapien
33	113.5	20.8	617	4	Q9Y4Z4	Q9Y4Z4 homo sapien
34	113.5	20.8	661	11	Q9TKS3	Q9TKS3 mus musculus
35	113.5	20.8	723	11	Q9KRS4	Q9KRS4 mus musculus
36	113.5	20.8	723	11	Q9WVH2	Q9WVH2 mus musculus
37	113.5	20.8	734	4	Q75112	Q75112 homo sapien
38	112.5	20.6	214	11	Q9CRA2	Q9CRA2 mus musculus
39	112.5	20.6	219	11	Q9Z250	Q9Z250 rattus norv
40	112.5	20.6	233	4	Q14910	Q14910 homo sapien
41	112.5	20.6	239	11	Q9QYNO1	Q9QYNO1 mus musculus
42	112.5	20.6	325	4	Q9H4L9	Q9H4L9 homo sapien
43	112.5	20.6	327	11	Q9KRS3	Q9KRS3 mus musculus
44	112.5	20.6	329	4	Q9BP29	Q9BP29 homo sapien
45	112.5	20.6	352	6	Q9CKU1	Q9CKU1 macaca fasc

RESULT 1
ID Q9QXN0 PRELIMINARY; PRT; 1986 AA.
AC Q9QXN0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PDZ DOMAIN ACTIN BINDING PROTEIN SHROOM.
GN SHRM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP MEDLINE=20055594; PubMed=10589677;
RA Hildebrand J.D., Soriano P.;
RT "Shroom, a PDZ domain-containing actin-binding protein, is required
for neural tube morphogenesis in mice."
RL Cell 99:485-497(1999).
DR EMBL; AF199421; AAF13269.1; -
DR HSSP; P29476; 10AV.
DR MGD; MGI:1351655; Shrm.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
SQ SEQUENCE 1986 AA; 215261 MW; 18C5EFCB772EABF CRC64;

Query Match 75.0%; Score 409; DB 11; Length 1986;
Best Local Similarity 75.5%; Pred. No. 4.3e-32;
Matches 80; Conservative 11; Mismatches 15; Indels 0; Gaps 0;

QY 2 MRRTEDEHFKPSATLNSNTAKRYIYLEAFLEGAPMGFTLKGLHGEPLISKVERGG 61
DB 1 MRRTEDEHFKPSATLNSNTAKRYIYLEAFLEGAPMGFTLKGLHGEPLISKVERGG 60
QY 62 KADTLSSKLAQDEYVHINEYTLSSSRKRAVSLVKGSKYTLRLVYR 107
DB 61 KADVSGLAQDEYVHINEYTLSSSRKRAVSLVKGSKYTLRLVYR 106

```

Matches 27; Conservative 13; Mismatches 10; Indels 1; Gaps 1;

QY 57 VEEGKADTLSSKLOAGDEVVHNEVTLSSSRKKAIVSLVKGSKYTLRLVVR 107
DB 1 IEDGGKA-ALSKMRTGDELVNINGTPLYGSRQEAALILKSGFRILKLIVR 50

RESULT 4
Q9H4T0 PRELIMINARY; PRT; 39 AA.
ID Q9H4T0
AC Q9H4T0;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE BA554P16.1 (KIAA1202) (FRAGMENT).
GN BA554P16.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL359272; CAC12948.1;
FT NON_TER 39
SQ SEQUENCE 39 AA; 4260 MW; 2773DAE613D25D3F CRC64;

Query Match 23.5%; Score 128; DB 4; Length 39;
Best Local Similarity 67.6%; Pred. No. 4.3e-06;
Matches 23; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 23 GRYIYLEAFLEGGAPWGFTLKGGLGHEGLIISK 56
DB 6 GSFQYVPVQLGGAPWGFTLKGGLGHEGLIISK 39

RESULT 5
Q9PU46 PRELIMINARY; PRT; 298 AA.
ID Q9PU46
AC Q9PU46;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SKELETAL MUSCLE ISOFORM (SKALP)
DE (FRAGMENT).
GN ALP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436131; Pubmed=10506181;
RA Ponies P., Macalima T., Beckerie M.C.;
RT "Purification and characterization of an alpha-actinin-binding PDZ-LIM
protein that is up-regulated during muscle differentiation.";
J. Biol. Chem. 274:29242-29250(1999).
RL EMBL; AJ249219; CAB53971.1;
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
FT NON_TER 298
SQ SEQUENCE 298 AA; 32094 MW; 27D6F24FC6141E61 CRC64;

Query Match 23.0%; Score 125.5; DB 13; Length 298;
Best Local Similarity 37.7%; Pred. No. 0.0001;
Matches 29; Conservative 16; Mismatches 29; Indels 3; Gaps 2;

RESULT 2
Q9ULL8 PRELIMINARY; PRT; 1502 AA.
ID Q9ULL8
AC Q9ULL8;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE KIAA1202 PROTEIN (FRAGMENT).
GN KIAA1202.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; Pubmed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033028; BA866516.1;
DR HSSP; P31016; IBE9.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 1502 AA; 165950 MW; DC050C9C69B78AD9 CRC64;

Query Match 48.2%; Score 262.5; DB 4; Length 1502;
Best Local Similarity 58.8%; Pred. No. 1.5e-17;
Matches 50; Conservative 17; Mismatches 17; Indels 1; Gaps 1;

QY 23 GRYIYLEAFLEGGAPWGFTLKGGLGHEGLIISKVEEGKADTLSSKLOAGDEVVHNEV 82
DB 10 GSFQYVPVQLGGAPWGFTLKGGLGHEGLIISKVEEGKADTLSSKLOAGDEVVHNEV 68

QY 83 TLSSSRKKAIVSLVKGSKYTLRLVVR 107
DB 69 PLYGSRQEAALILKSGFRILKLIVR 93

RESULT 3
Q9H4G2 PRELIMINARY; PRT; 1454 AA.
ID Q9H4G2
AC Q9H4G2;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE BAI19E20.1 (KIAA1202 PROTEIN) (FRAGMENT).
GN BAI19E20.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson S.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121865; CAC08472.1;
DR InterPro; IPR001478; PDZ.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 1454 AA; 160615 MW; 7931571D7DA80DE8 CRC64;

Query Match 24.7%; Score 134.5; DB 4; Length 1454;
Best Local Similarity 52.9%; Pred. No. 0.0001;

```

OY 32 LEEGAPWFTLKGLGHEGEPILISKVEEGKADPTLSSKIQAGDEVHINEV-TLSSSRKE 90
 DB 7 LPPAPAPGFTLSSGIFNCPILITRITPGSKAST--ANLCPGDIYAINGLSTENWTHND 64
 OY 91 AVSLVKGSKYTLRLVYR 107
 DB 65 AOKRIKAHQLSLIRIK 81
 RESULT 6
 ID Q9HAP6 PRELIMINARY; PRT; 207 AA.
 AC Q9HAP6
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE LIN-7B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 RA Olsen O., Liu H., Merot J., Wellington P.A.;
 RT "Basolateral Membrane Expression of the Kir 2.3 Channel is Coordinated
 RT by a PDZ Interaction with Lin-7/ASK.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF311862; AAC34117.1;
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ. 1.
 DR SMART: SM00228; PDZ. 1.
 DR PROSITE: PS50106; PDZ. 1.
 DR SEQUENCE 207 AA; 22896 MW; 63189D82706B9B00 CRC64;
 SO
 Query Match 22.8%; Score 124.5; DB 4; Length 207;
 Best Local Similarity 36.0%; Pred. No. 8.3e-05;
 Matches 36; Conservative 21; Mismatches 34; Indels 9; Gaps 4;
 OY 13 ATLNSNTATKG---RYILEAFLEGAPWFTLKGLGHEGEPILISKVEEGKADPTLSS 68
 DB 77 ATVAAPTASGSHAPRVVLPKPTDEG--LGNIMGKQNSPITISVITPGVADR-HG 132
 OY 69 KIQAGDEVHINEVTLSSSRKE-AVSLVKGSKYTLRLVYR 107
 DB 133 GLKRGDQLSVNGSVGEHHEKAVELLLKAAQGSYKLVYR 172
 RESULT 7
 ID Q9Z252 PRELIMINARY; PRT; 207 AA.
 AC Q9Z252
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE LIN-7-A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Taki Y., Hata Y., Deguchi M., Ide N., Hirao K., Yao I., Nishiohara H.,
 RT "Isolation and Characterization of Mammalian Homologues of
 RT Caenorhabditis elegans Lin-7: Localization at Cell-Cell Junction.";
 RL OncoGene 0:0-0(1999).
 DR EMBL: AF090133; AAC78072.1;
 DR HSSP: Q12923; 3PDZ.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ. 1.
 DR SMART: SM00228; PDZ. 1.
 DR SEQUENCE 207 AA; 22914 MW; 6C6B6754B8C89F03 CRC64;
 SO

DR PROSITE: PS50106; PDZ. 1.
 SQ SEQUENCE 207 AA; 22900 MW; 6C6B6754B8C89B13 CRC64;
 Query Match 22.7%; Score 123.5; DB 11; Length 207;
 Best Local Similarity 36.0%; Pred. No. 0.0001;
 Matches 36; Conservative 20; Mismatches 35; Indels 9; Gaps 4;
 OY 13 ATLNSNTATKG---RYILEAFLEGAPWFTLKGLGHEGEPILISKVEEGKADPTLSS 68
 DB 77 ATVAAPTASGSHAPRVVLPKPTDEG--LGNIMGKQNSPITISVITPGVADR-HG 132
 OY 69 KIQAGDEVHINEVTLSSSRKE-AVSLVKGSKYTLRLVYR 107
 DB 133 GLKRGDQLSVNGSVGEHHEKAVELLLKAAQGSYKLVYR 172
 RESULT 8
 ID O88951 PRELIMINARY; PRT; 207 AA.
 AC O88951
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE VELI 2
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98424246; PubMed=9753324;
 RA Butz S., Okamoto M., Sudhof T.C.;
 RT "A tripartite protein complex with the potential to couple synaptic
 RT vesicle exocytosis to cell adhesion in brain.";
 RL Cell 94:773-782(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99274724; PubMed=10341223;
 RA Jo K., Derin R., Li M., Bredt D.S.;
 RT "Characterization of MLVs/Velis-1, -2, and -3: a family of mammalian
 RT Lin-7 homologs enriched at brain synapses in association with the
 RT postsynaptic density-95/NMDA receptor postsynaptic complex.";
 RL J. Neurosci. 19:4189-4199(1999).
 DR EMBL: AF087694; AAC78482.1;
 DR HSSP: AF173082; AAP48501.1;
 DR MGD: MGI:130858; Vel12.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ. 1.
 DR SMART: SM00228; PDZ. 1.
 DR PROSITE: PS50106; PDZ. 1.
 DR SEQUENCE 207 AA; 22914 MW; 6C6B6754B8C89F03 CRC64;
 SO
 Query Match 22.7%; Score 123.5; DB 11; Length 207;
 Best Local Similarity 36.0%; Pred. No. 0.0001;
 Matches 36; Conservative 20; Mismatches 35; Indels 9; Gaps 4;
 OY 13 ATLNSNTATKG---RYILEAFLEGAPWFTLKGLGHEGEPILISKVEEGKADPTLSS 68
 DB 77 ATVAAPTASGSHAPRVVLPKPTDEG--LGNIMGKQNSPITISVITPGVADR-HG 132
 OY 69 KIQAGDEVHINEVTLSSSRKE-AVSLVKGSKYTLRLVYR 107
 DB 133 GLKRGDQLSVNGSVGEHHEKAVELLLKAAQGSYKLVYR 172
 RESULT 9
 ID Q9D201 PRELIMINARY; PRT; 234 AA.
 AC Q9D201

DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE VTEBRATE HOMOLOGY OF C. ELEGANS LIN-7 TYPE 2.
 GN VELI2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arawaka T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleisemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK019299; BAB31655.1;
 DR MGD; MGI:1330858; Veli2.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 SQ SEQUENCE 234 AA; 25732 MW; 76919D8B1C29E53E CRC64;

 Query Match 22.7%; Score 123.5; DB 11; Length 234;
 Best Local Similarity 36.0%; Pred. No. 0.00012;
 Matches 36; Conservative 20; Mismatches 35; Indels 9; Gaps 4;

 QY 13 ATLNNTATKG----RYIYLEAFLEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSS 68
 DB 77 ATVAFTASEGHAHPVVELPKTDG---LGFNIMGKQEQNSPIYISRVIPGVADR-HG 132
 QY 69 KIQAGDEVVHINEVTLSSSRKE-AVSLVKGSKYKTLRLVVR 107
 DB 133 GLKRGDQLLSVNGSVGEHHEKAVELLKAAQGSVKLVVR 172

 RESULT 10
 ID Q9PU47 PRELIMINARY; PRT; 315 AA.
 AC Q9PU47;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SMOOTH MUSCLE ISOFORM (SMALP).
 GN ALP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99436131; PubMed=10506181;
 RA Pomies P., Macalma T., Beckerle M.C.;

RT "Purification and characterization of an alpha-actinin-binding PDZ-LIM
 RT protein that is up-regulated during muscle differentiation."
 RL J. Biol. Chem. 274:29242-29250(1999).
 CC -1- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
 CC IONS.
 CC EMBL; AJ249218; CAB53970.1;
 DR InterPro; IPR001781; LIM.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00412; LIM; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR ProDom; PD000094; LIM; 1.
 DR SMART; SM00132; LIM; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR LIM motif; Metal-binding; Zinc.
 SQ SEQUENCE 315 AA; 34365 MW; 86414A80AB9BCA35 CRC64;

Query Match 22.7%; Score 123.5; DB 13; Length 315;
 Best Local Similarity 39.2%; Pred. No. 0.00018;
 Matches 29; Conservative 14; Mismatches 28; Indels 3; Gaps 2;

 QY 32 LEGAPMGFTLKGLEHGEPLIISKVEGGKADTLSSKLQAGDEVVHINEV-TLSSSRKE 90
 DB 7 LPGAAPMGFRLSGIDFNQPLIITRITPGSKAST--ANLCPGDIIVAINGLSTENMTHND 64

 QY 91 AVSLVKGSKYKTLRL 104
 DB 65 AQERIKAAHQLSL 78

RESULT 11
 ID 017458 PRELIMINARY; PRT; 171 AA.
 AC 017458;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE HYPOTHETICAL 19.0 KDA PROTEIN (FRAGMENT).
 DE Schistosoma mansoni (Blood fluke).
 OS Schistosoma.
 OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
 OC Trematoda; Digenea; Strigeidae; Schistosomatidae; Schistosomatidae;
 OC Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PUERTO RICAN;
 RA Harrop R., Wilson R.A.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF030965; AAB86566.1;
 DR HSSP; P31016; 1BFE.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 171 AA; 19000 MW; 69685F4108598063 CRC64;

Query Match 22.6%; Score 123; DB 5; Length 171;
 Best Local Similarity 34.7%; Pred. No. 9.2e-05;
 Matches 41; Conservative 20; Mismatches 39; Indels 18; Gaps 6;

QY 4 TTEDEH----KPSATLNSN---TATKG---RYIYLEAFLEGGAPMGFTLKGLEHGE 50
 DB 34 TTVYDINGSEEVKASATAKATAFAAFAEGHAHPRIELPKTNKG---LGFNVYMGKEQNS 90
 QY 51 PLITISKVEGGKADTLSSKLQAGDEVVHINEVTLSSSRKE-AVSLVKGSKYKTLRLVVR 107
 DB 91 PIYISRMXPGGVADR-HGGLKRGDQLLSVNGISVESEHRAVELLKLAAQGVTKLVVR 147

[illegible][illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:54:52 ; Search time 56.12 Seconds
(Without alignments)
250.783 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MARRTEDFKPSATLNSNTA.....CIYPROHNTCRNFKRAYLSR 190

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq-1101.*
1: /SID52/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1029	100.0	190	22	AAU00031
2	545	53.0	108	22	AAU00032
3	541	52.6	133	22	AAU00033
4	132.5	12.9	150	21	AA03364
5	129	12.5	113	21	AA02039
6	128.5	12.5	207	22	AA03655
7	126.5	12.3	403	20	AA07095
8	126.5	12.3	406	19	AA07036
9	126.5	12.3	481	20	AA07098
10	126.5	12.3	521	20	AA07099
11	126.5	12.3	652	20	AA07094

12	126.5	12.3	652	20	AA07050	Renal cancer assoc
13	125.5	12.2	329	19	AA058396	Homo sapiens HLIM-
14	124.5	12.1	153	21	AA036475	Human LMP-3 (HLMP-
15	124.5	12.1	223	20	AA097845	Human truncated bo
16	124.5	12.1	223	21	AA036473	Human LMP-1 amino
17	124.5	12.1	457	20	AA097844	Human bone mineral
18	124.5	12.1	457	21	AA036472	Human LIM mineral
19	122.5	11.9	316	19	AA058397	Homo sapiens HLIM-
20	122.5	11.9	316	19	AA058397	Human prostate can
21	122.5	11.9	913	18	AA012522	Protein tyrosine p
22	122.5	11.9	913	18	AA012522	Human PTPB1 protei
23	122.5	11.9	127	20	AA025156	Human 5' EST seque
24	121.5	11.8	454	17	AA088804	Human insulin rece
25	121.5	11.8	469	22	AA040870	Human polypeptide
26	121.5	11.8	632	21	AA066889	Membrane-bound pro
27	121.5	11.8	632	22	AA039084	Human polypeptide
28	121.5	11.8	632	22	AA087545	Human PRO1136, Ho
29	121.5	11.8	632	22	AA065212	Human PRO1136, UNQ
30	119.5	11.6	86	22	AA055840	PDZ encoded domain
31	119.5	11.6	86	22	AA057631	Human LIM protein
32	119.5	11.6	86	22	AA058041	Human LIM protein
33	119.5	11.6	423	21	AA036474	Human LMP-2 (HLMP-
34	119.5	11.6	457	20	AA097843	Rat bone mineralis
35	119.5	11.6	457	21	AA036471	Human protein sequ
36	119.5	11.6	197	22	AA093600	Peptide #6916 enco
37	117	11.4	37	22	AA032879	Human protein sequ
38	117	11.4	352	22	AA095452	Neuron-associated
39	116.5	11.3	632	21	AA01385	Human cell junctio
40	116	11.3	233	20	AA029978	Human cell junctio
41	116	11.3	233	22	AA07127	Human colon cancer
42	115.5	11.2	227	22	AA053276	Human colon cancer
43	115.5	11.2	227	22	AA05128	Human PDZ protein
44	115	11.2	181	22	AA064378	Human PDZ protein
45	112	10.9	817	19	AA048101	Human discs large

ALIGNMENTS

RESULT 1	AAU00031	standard; Protein; 190 AA.
ID	AAU00031	
XX	AAU00031	
AC	11-MAY-2001 (first entry)	
XX		
DE	Novel human protein, NHP#1.	
XX		
KW	Human; novel human protein; NHP#1; gene therapy; drug screening;	
KW	obesity; high blood pressure.	
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Misc-feature	1
FT		/note="Absent if the true start codon is
FT		nucleotides 4-6 of sequence in AA000039, the authors
FT		are uncertain which is the true start codon"
XX		
XX	WO200114422-A1.	
XX		
PD	01-MAR-2001.	
XX		
XX	18-AUG-2000; 2000MO-US22815.	
PF		
XX	24-AUG-1999; 99US-0150511.	
PR		
XX	(LEXI-) LEXICON GENETICS INC.	
PA		
PI	Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;	
XX		
DR	WPI: 2001-218430/22.	

DR N-PSDB; AAS00031.

XX Novel human polynucleotides isolated from human mammary gland cDNA

PT library, encodes novel human proteins which are useful in diagnosis,

PT drug screening, clinical trial monitoring or treating behavioural

PT disorders

XX

XX Claim 2; Page 24-25; 29pp; English.

XX

XX The sequence is a novel human protein, NHP#1, which shares

CC structural motifs with human APXL protein. Nucleotide constructs

CC encoding functional NHPs are used in gene therapy approaches for the

CC modulation of NHP expression. NHP oligonucleotides can be used as

CC hybridisation probes for screening libraries and assessing NHP gene

CC expression patterns. Also, labeled NHP nucleotide probes can be used to

CC screen a human genomic library. The NHP nucleotide sequences are also

CC useful in drug screening techniques for treating symptomatic or

CC phenotypic manifestations of perturbing the normal function of NHP in the

CC body. Examples of such manifestations may include obesity and high blood

CC pressure.

XX Sequence 190 AA;

Query Match 100.0%; Score 1029; DB 22; Length 190;

Best Local Similarity 100.0%; Pred. No. 1.8e-107;

Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMRTEDEPHKPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGLEHGPPLIISKVEEG 60

DB 1 mmrttedfhkpsatlnsntatkgriyyleaflegapwgftlkgglehgppliskveeg 60

QY 61 GKADTLSSKLQAGDEVVHINEVTLSRRKEAVSLVKGYSKTLRLVVRSLPPVTVSLEFD 120

DB 61 gkadtlssklqagdevvhinevtlssrkeavslvkgysktrlrvslsppvtslefd 120

QY 121 PQHPQMPPTRTSFSVSTADGRHESCRRPWPVKWSPRPTTAAARWPKGCIYPTQNTC 180

DB 121 pqhpmprtrtsfsvstadgrhescrrpwpvkwwsprpttwaarwpkgciyptqntc 180

QY 181 RNFKRAYLSR 190

DB 181 rnfkraylsr 190

RESULT 2

AAU00032

ID AAU00032 standard; Protein; 108 AA.

XX AC

XX AAU00032;

DT 11-MAY-2001 (first entry)

XX

XX Novel human protein, NHP#2.

DE Human; novel human protein; NHP#2; gene therapy; drug screening;

KW obesity; high blood pressure.

XX

OS Homo sapiens.

XX

XX Key Location/Qualifiers

FT Misc_feature 1

FT /note= "Absent if the true start codon is

FT nucleotides 4-6 of sequence in AAS00040. the authors

FT are uncertain which is the true start codon"

XX

XX WO200114422-A1.

PN

XX

XX 01-MAR-2001.

PD

XX

XX 18-AUG-2000; 2000WO-US22815.

PF

XX

XX 24-AUG-1999; 99US-0150511.

PR

XX (LEXI-) LEXICON GENETICS INC.

XX

XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;

XX WPI; 2001-218430/22.

DR N-PSDB; AAS00040.

XX

XX Novel human polynucleotides isolated from human mammary gland cDNA

PT library, encodes novel human proteins which are useful in diagnosis,

PT drug screening, clinical trial monitoring or treating behavioural

PT disorders

XX

XX Claim 4; Page 25; 29pp; English.

XX

XX The sequence is a novel human protein, NHP#2, which shares

CC structural motifs with human APXL protein. Nucleotide constructs

CC encoding functional NHPs are used in gene therapy approaches for the

CC modulation of NHP expression. NHP oligonucleotides can be used as

CC hybridisation probes for screening libraries and assessing NHP gene

CC expression patterns. Also, labeled NHP nucleotide probes can be used to

CC screen a human genomic library. The NHP nucleotide sequences are also

CC useful in drug screening techniques for treating symptomatic or

CC phenotypic manifestations of perturbing the normal function of NHP in the

CC body. Examples of such manifestations may include obesity and high blood

CC pressure.

XX Sequence 108 AA;

Query Match 53.0%; Score 545; DB 22; Length 108;

Best Local Similarity 100.0%; Pred. No. 2e-53;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMRTEDEPHKPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGLEHGPPLIISKVEEG 60

DB 1 mmrttedfhkpsatlnsntatkgriyyleaflegapwgftlkgglehgppliskveeg 60

QY 61 GKADTLSSKLQAGDEVVHINEVTLSRRKEAVSLVKGYSKTLRLVVRSLPPVTVSLEFD 108

DB 61 gkadtlssklqagdevvhinevtlssrkeavslvkgysktrlrvslsppvtslefd 108

RESULT 3

AAU00033

ID AAU00033 standard; Protein; 133 AA.

XX AC

XX AAU00033;

DT 11-MAY-2001 (first entry)

XX

XX Novel human protein, NHP#3.

DE Human; novel human protein; NHP#3; gene therapy; drug screening;

KW obesity; high blood pressure.

XX

OS Homo sapiens.

XX

XX Key Location/Qualifiers

FT Misc_feature 1

FT /note= "Absent if the true start codon is

FT nucleotides 4-6 of sequence in AAS00041. the authors

FT are uncertain which is the true start codon"

XX

XX WO200114422-A1.

PN

XX

XX 01-MAR-2001.

PD

XX

XX 18-AUG-2000; 2000WO-US22815.

PF

XX

XX 24-AUG-1999; 99US-0150511.

PR

XX (LEXI-) LEXICON GENETICS INC.

XX PI Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
 XX DR WPI: 2001-218430/22.
 DR N-PSDB: AAS00041.
 XX PT Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX PS Claim 6; Page 25-26; 29pp; English.
 CC The sequence is a novel human protein, NHP#3, which shares
 CC structural motifs with human APXL protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide sequences are also
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 CC CC
 CC CC
 XX Sequence 133 AA;
 SQ
 Query Match 52.6%; Score 541; DB 22; Length 133;
 Best Local Similarity 100.0%; Pred. No. 7.8e-53;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 NMRTEDFKPATLNSNTATGRITYLEAFLEGAAPWGFITLKGLEHPEPLITSVEEG 60
 DB 1 mrttedfhpaltlnsatatgryyleaflegapwgfittkglhpeplitsveeg 60
 OY 61 GRADTSSKLGAGDEVYHINVTSSSRKREAVSLVGSYKTLIVVR 107
 DB 61 gxadtlssklgagdevyhinevtlsssrkxavslvgsyktlilivr 107
 RESULT 4
 AAB43364
 ID AAB43364 standard; Protein; 150 AA.
 XX AAB43364;
 AC
 XX
 XX
 DT 08-FEB-2001 (first entry)
 DT
 XX
 DE Human ORFX ORF3128 polypeptide sequence SEQ ID NO:6256.
 KW Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
 KW vulnerrary; antiproliferic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 XX Homo sapiens.
 XX
 XX MO200058473-A2.
 PN
 XX
 XX 05-OCT-2000.
 PD
 XX
 XX 31-MAR-2000; 2000MO-US08621.
 PF

XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 XX Shinkets RA, Leach M;
 PI
 XX WPI: 2000-602362/57.
 DR N-PSDB: AAC77573.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease
 PT
 XX Claim 11; Page 5443; 5507pp; English.
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnerrary;
 CC antiproliferic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antifungal; antirheumatic; antihypertoid;
 CC antihypertoid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC CC
 CC CC
 XX Sequence 150 AA;
 SQ
 Query Match 12.9%; Score 132.5; DB 21; Length 150;
 Best Local Similarity 32.6%; Pred. No. 7.9e-07;
 Matches 46; Conservative 27; Mismatches 53; Indels 15; Gaps 7;
 OY 10 KPSATLNSNTATKG---RYILEAFLEGAAPWGFITLKGLEHPEPLITSVEEGKADT 65
 DB 17 kkatvaatlaseghahprvvelpkrtdg---lglfimgskqnspllyslvpgvad 73
 OY 66 LSKLQAGDEVYHINVTSSSRKE-ASVLVGSYKTLIVVRSLSPPTVSLF--FDPQ 122
 DB 74 -hgltrggqllsvnvsvsgeqheavellkaagsvklivr-ytpryleamearfelm 131
 OY 123 HPOKMPRTRTSFSVSATDGR 143
 DB 132 rsar---rrqghsystsleer 149
 RESULT 5
 AAG02039
 ID AAG02039 standard; Protein; 113 AA.
 XX AAG02039;
 AC
 XX
 XX 06-OCT-2000 (first entry)
 DT
 XX
 XX Human secreted protein, SEQ ID NO: 6120.
 DE
 XX
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 KW


```

OY 13 ATLNSTARK---RYILEAFLESGAPMGFTLKGLHEHPELLISKEYEFGKADTLSS 68
DB 77 atvaafaseghphrvvelpktdge---lgfnimgkqsplyisrylpgvadtr-hg 132
OY 69 KLOAGDEVHINVTLSSSRKE-AVSLVKGSKYKTLRLVRSLSPPVTASLE--FDPQHPQ 125
DB 133 gktrgqdlisvngsvgeqehakavellikaagsvklvtr-yprvrlcearfekmrae 191
OY 126 RMPRTTSFSVSTADGR 143
DB 192 r---rrqbgqsyssleer 206

RESULT 7
AAV07095 standard; Protein; 403 AA.
XX AC AAV07095;
XX DT 02-JUL-1999 (first entry)
XX DE Colon cancer associated antigen precursor sequence.
XX KM Cancer associated antigen; diagnosis; research; treatment; human;
XX KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX KM prostate cancer.
XX OS Homo sapiens.
XX PN WC9904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98MO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Opat Y, Old LJ;
XX PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
XX PI Tureci O;
XX WPI: 1999-132448/11.
XX PS New isolated cancer associated nucleic acids and polypeptides -
XX PT isolated using sera from cancer patients, used to develop products
XX PT for the diagnosis, monitoring or treatment of cancers
XX PS Disclosure: Page 659-660; 787pp; English.
XX PS The invention relates to a method for diagnosing a disorder characterised
XX CC by expression of a human cancer associated antigen precursor coded for by
XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX CC biological sample isolated from a subject with an agent that specifically
XX CC binds to the NAM, an expression product or a fragment of an expression
XX CC product complexed with an RNA molecule; and (b) determining the
XX CC interaction between the agent and the NAM or the expression product as a
XX CC determination of the disorder. The products and methods can be used in
XX CC the diagnosis, monitoring, research, or treatment of conditions
XX CC characterised by the expression of various cancer associated antigens.
XX CC The invention provides nucleic acid sequences and encoded polypeptides
XX CC which are cancer associated antigen precursors expressed in human breast
XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX CC lung cancer.
XX SQ Sequence 403 AA;

```

```

Query Match 12.3%; Score 126.5; DB 20; Length 403;
Best Local Similarity 35.6%; Pred. No. 1.6e-05;
Matches 31; Conservative 24; Mismatches 25; Indels 7; Gaps 4;

OY 39 GFTLKGLHEHPELLISKEYEFGKADTLSSKIQADDEVHINEVTLSS-SRKAVALYK 97
DB 99 glsvrgglfegcglflshllykqgdsvq--lqvdeylrnygyslscneevlnllr- 155
OY 98 SYKTLRLVRSLSPPVTASLEFDPQHP 124
DB 156 tkktvsikvrhlyg---lprvkspsdep 179

RESULT 8
AAV070326 standard; Protein; 406 AA.
XX AC AAV070326;
XX DT 21-DEC-1998 (first entry)
XX DE Secreted protein DU123_1.
XX DE Secreted protein; DU123_1; human.
XX KM Homo sapiens.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT MISC-difference 390
XX FT /note="encoded by AAV"
XX PN WC9838209-A2.
XX PD 03-SEP-1998.
XX PF 25-FEB-1998; 98MO-US03697.
XX PR 24-FEB-1998; 98US-0028724.
XX PR 26-FEB-1997; 97US-0805819.
XX PA (GEMV ) GENETICS INST INC.
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX PI Racie LA, Spaulding V, Treacy M;
XX DR WPI: 1998-481139/41.
XX DR N-PSDB: AAV33198.
XX PS New isolated polynucleotide(s) and encoded polypeptide(s) -
XX PT obtained from human foetal kidney, adult colon, adult brain, foetal
XX PT brain and placenta cDNA libraries.
XX PS Claim 33; Page 80-81; 103pp; English.
XX PS This is the amino acid sequence of novel human secreted protein
XX CC DU123_1, as deduced from a full-length cDNA clone (see AAV33198)
XX CC obtained from a human foetal brain cDNA library. Database
XX CC sequencing revealed some similarity between DU123_1 and some known
XX CC sequences. The invention provides new isolated polynucleotides
XX CC (see AAV33190-99), from human foetal kidney, adult colon, adult brain,
XX CC foetal brain and placenta cDNA libraries, that code for secreted
XX CC proteins (see AAV70319-27). The clones can be used for recombinant
XX CC production of the polypeptides, which may have activities such as
XX CC e.g. nutritional activity, immunostimulant or immunosuppressive,
XX CC differentiation activity, regulating activity, tissue growth activity,
XX CC haematopoiesis regulating activity, chemokine activity, activin
XX CC or inhibin activity, chemotactic or chemokinetic activity,
XX CC haemostatic and thrombolytic activity, receptor/ligand activity,
XX CC antiinflammatory activity, cadherin/tumour invasion suppressor
XX CC activity, tumour inhibition activity, or other activities.
XX SQ Sequence 406 AA;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:56:17 ; Search time 29.66 Seconds
(without alignments)
144.155 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MMRTEDEHRSATLNSMTA.....CIYPTQHTCRNFRAYLSR 190

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	12.3	327	2	US-08-739-485-10 Sequence 10, Appl
2	125.5	12.2	329	2	US-08-739-485-3 Sequence 3, Appl
3	124.5	12.1	223	4	US-09-124-238A-14 Sequence 34, Appl
4	124.5	12.1	457	4	US-09-124-238A-10 Sequence 10, Appl
5	122.5	11.9	316	2	US-08-739-485-5 Sequence 5, Appl
6	121.5	11.8	330	2	US-08-739-485-11 Sequence 11, Appl
7	121.5	11.8	454	1	US-08-166-316-2 Sequence 1, Appl
8	119.5	11.6	457	4	US-09-124-238A-1 Sequence 1, Appl
9	117	11.4	75	4	US-09-100-804-21 Sequence 21, Appl
10	116	11.3	233	2	US-09-151-611-1 Sequence 1, Appl
11	116	11.3	233	2	US-09-370-102-1 Sequence 1, Appl
12	110	10.7	297	4	US-09-151-611-3 Sequence 3, Appl
13	110	10.7	297	4	US-09-370-102-3 Sequence 3, Appl
14	108	10.5	505	1	US-08-123-161A-14 Sequence 14, Appl
15	108	10.5	505	1	US-08-483-278-14 Sequence 14, Appl
16	103.5	10.1	73	3	US-09-100-804-28 Sequence 28, Appl
17	103.5	10.1	498	3	US-09-045-632-30 Sequence 30, Appl
18	100.5	9.8	541	3	US-09-045-632-36 Sequence 36, Appl
19	100.5	9.8	599	3	US-09-045-632-28 Sequence 28, Appl
20	100.5	9.8	642	3	US-09-045-632-35 Sequence 35, Appl
21	100.5	9.8	818	3	US-09-045-632-25 Sequence 25, Appl
22	100.5	9.8	861	3	US-09-045-632-34 Sequence 34, Appl
23	100.5	9.8	918	3	US-09-045-632-21 Sequence 21, Appl
24	100.5	9.8	961	3	US-09-045-632-13 Sequence 13, Appl
25	100.5	9.8	1018	3	US-09-045-632-16 Sequence 16, Appl
26	100.5	9.8	1061	3	US-09-045-632-32 Sequence 32, Appl
27	100.5	9.8	1112	3	US-09-045-632-2 Sequence 2, Appl

28	100.5	9.8	1112	3	US-09-045-632-3	Sequence 3, Appl
29	95	9.2	182	3	US-09-045-632-29	Sequence 29, Appl
30	95	9.2	283	3	US-09-045-632-27	Sequence 27, Appl
31	95	9.2	502	3	US-09-045-632-24	Sequence 24, Appl
32	95	9.2	602	3	US-09-045-632-20	Sequence 20, Appl
33	95	9.2	702	3	US-09-045-632-15	Sequence 15, Appl
34	93.5	9.1	80	5	US-08-545-860D-54	Sequence 54, Appl
35	93.5	9.1	80	5	PCT-US94-04496-54	Sequence 54, Appl
36	93.5	9.1	398	3	US-09-045-632-31	Sequence 31, Appl
37	93.5	9.1	441	3	US-09-045-632-49	Sequence 49, Appl
38	93	9.0	1050	3	US-09-045-632-50	Sequence 50, Appl
39	93	9.0	1050	3	US-09-045-632-20	Sequence 25, Appl
40	92.5	9.0	73	3	US-09-100-804-25	Sequence 34, Appl
41	92.5	9.0	74	3	US-09-100-804-34	Sequence 11, Appl
42	91	8.8	198	3	US-09-045-632-11	Sequence 11, Appl
43	91	8.8	284	3	US-09-045-632-12	Sequence 12, Appl
44	91	8.8	507	3	US-09-045-632-13	Sequence 13, Appl
45	91	8.8	604	3	US-09-045-632-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-739-485-10
Sequence 10, Application US/08739485
Patent No. 5663898
GENERAL INFORMATION:
APPLICANT: Goli, Surya K.
APPLICANT: Hallman, Jennifer L.
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/739,485
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0142 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1020151
US-08-739-485-10
Query Match 12.3%; Score 127; DB 2; Length 327;

Db 122 GISLEDNR 129

RESULT 7

US-08-166-316-2

Sequence 2, Application US/08166316

Patent No. 5504192

GENERAL INFORMATION:

APPLICANT: Gill, Gordon N.

APPLICANT: Wu, Rui-Yun

TITLE OF INVENTION: ENDOCYTIC CODE BINDING PROTEIN

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Spensley Horn Jubas & Lubitz

STREET: 1880 Century Park East, Suite 500

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/166.316

FILING DATE: 13-DEC-1993

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Wetherell, Jr., Ph.D., John R.

REGISTRATION NUMBER: 31,678

REFERENCE/DOCKET NUMBER: PD-3204

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 455-5100

TELEFAX: (619) 455-5110

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 454 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-166-316-2

Query Match 11.8%; Score 121.5; DB 1; Length 454;

Best Local Similarity 24.4%; Pred. No. 1.7e-05;

Matches 48; Conservative 24; Mismatches 70; Indels 55; Gaps 6;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90

Db 8 LEGPAPWGFRLOGGKDFNVPLSISRLTPGGKA--AQSGVGDWVLSIDGENAGSLTHIE 64

QY 91 AVSLVK-----GSYKT-----LRLVVRSLSPVTV 115

Db 65 AQNKIRACGEAPGPGQPGAGSEQTAEGLRPRGPSAYTFAPSVSLNKTAAQPGPGAD 124

QY 116 SLEFPDQHPQR--MPRTTSTFSVSTADGRHWSRPPVWKSPPRTWAARWPKGCIY 173

Db 125 SPPOONGQPLRLVLPDASKQRLMENTED-----WRPRQASRVPSASLPT 170

QY 174 PTOHNTCRNEKRAYLSR 190

Db 171 SQAPSSCKTPDEHLKK 187

RESULT 8

US-09-124-238A-1

Sequence 1, Application US/09124238A

Patent No. 6300127

GENERAL INFORMATION:

APPLICANT: Hair, Gregory A.

APPLICANT: Boden, Scott D.

TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90

Db 8 LEGPAPWGFRLOGGKDFNVPLSISRLTPGGKA--AQSGVGDWVLSIDGENAGSLTHIE 64

QY 91 AVSLVK-----GSYKT-----LRLVVRSLSPVTV 115

Db 65 AQNKIRACGEAPGPGQPGAGSEQTAEGLRPRGPSAYTFAPSVSLNKTAAQPGPGAD 124

QY 116 SLEFPDQHPQR--MPRTTSTFSVSTADGRHWSRPPVWKSPPRTWAARWPKGCIY 173

Db 125 SPPOONGQPLRLVLPDASKQRLMENTED-----WRPRQASRVPSASLPT 170

QY 174 PTOHNTCRNEKRAYLSR 190

Db 171 SQAPSSCKTPDEHLKK 187

RESULT 9

US-09-100-804-21

Sequence 21, Application US/09100804

Patent No. 6066472

GENERAL INFORMATION:

APPLICANT: GONEZ, LEONEL JORGE

APPLICANT: SARAS, JAN

APPLICANT: CLAESON-WELSH, LENA

APPLICANT: HELDIN, CARL-HENRIK

TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL

TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN

TITLE OF INVENTION: TYROSINE PHOSPHATASES

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/100.804

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/596.291

FILING DATE: 09-AUG-1996

APPLICATION NUMBER: US 08/115.573

FILING DATE: 01-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/09943

FILING DATE: 01-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90

Db 8 LEGPAPWGFRLOGGKDFNVPLSISRLTPGGKA--AQSGVGDWVLSIDGENAGSLTHIE 65

QY 91 AVSLVKGYKTLRLVVRSLSPVTVSLEFDPQH---PORMPPRTTSTFSVSTADGRHWS 147

Db 66 AQNKIRACGERLSGLSRAQP-----AQSKPKALTPPADPPRYTFAPSASLNKTARPPG 120

QY 148 CRPPW-----VKWWSRPP 160

Db 121 APPDTSALSONGOLLROLVPDASKQRLMENTEDWRPRP 159

Query Match 11.6%; Score 119.5; DB 4; Length 457;

Best Local Similarity 27.7%; Pred. No. 2.9e-05;

Matches 44; Conservative 18; Mismatches 60; Indels 37; Gaps 5;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90

Db 8 LEGPAPWGFRLOGGKDFNVPLSISRLTPGGKA--AQSGVGDWVLSIDGENAGSLTHIE 65

QY 91 AVSLVKGYKTLRLVVRSLSPVTVSLEFDPQH---PORMPPRTTSTFSVSTADGRHWS 147

Db 66 AQNKIRACGERLSGLSRAQP-----AQSKPKALTPPADPPRYTFAPSASLNKTARPPG 120

QY 148 CRPPW-----VKWWSRPP 160

Db 121 APPDTSALSONGOLLROLVPDASKQRLMENTEDWRPRP 159

TITLE OF INVENTION: Expression Systems

FILE REFERENCE: 06148.0115

CURRENT APPLICATION NUMBER: US/09/124.238A

CURRENT FILING DATE: 1998-07-29

PRIOR APPLICATION NUMBER: 60/054,219

PRIOR FILING DATE: 1997-07-30

PRIOR APPLICATION NUMBER: 60/080,407

PRIOR FILING DATE: 1998-04-02

NUMBER OF SEQ ID NOS: 36

SOFTWARE: MS Word

SEQ ID NO 1

LENGTH: 457

TYPE: PRT

ORGANISM: Rattus norvegicus

US-09-124-238A-1

REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: LC461/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-100-804-21

Query Match 11.4%; Score 117; DB 3; Length 75;
Best Local Similarity 38.8%; Pred. No. 3.8e-06;
Matches 26; Conservative 12; Mismatches 27; Indels 2; Gaps 2;

QY 33 EGAPNGFTLKGLEHGEPLIISKVEEGKADTLSSKIQAGDEVHINEVTLSS-SRKEA 91
DB 2 EDKRP-GEFLKGVDOKNPLVSRINSSPADYICPKLNEGDOVLINGRDISEHTDQV 60

QY 92 VSLVKG 98
DB 61 VMFKAS 67

RESULT 10
US-09-151-611-1
Sequence 1, Application US/09151611
Patent No. 5958731
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Au-Young, Janice
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
FILE REFERENCE: PF-0599 US
CURRENT APPLICATION NUMBER: US/09/151,611
CURRENT FILING DATE: 1998-09-11
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 233
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 1974337
US-09-151-611-1

Query Match 11.3%; Score 116; DB 2; Length 233;
Best Local Similarity 32.4%; Pred. No. 2.7e-05;
Matches 36; Conservative 24; Mismatches 41; Indels 10; Gaps 5;

QY 13 ATLNSNTATKG---RYYLEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLSS 68
DB 92 ATVAAFASSEGHSHRPVELPKTDEG---LGFNVGKGEONSPYISRIIPGVADR-HG 147
QY 69 KIQAGDEVHINEVTLSSRKE-AVSLVKGSYKTLRLVYRSLSPPVTSLE 118
DB 148 GLKRGDQLSLVNGSVGEHHEKAVELLKAKADSVKLVVR-VTPKYLEEME 197

RESULT 11
US-09-370-102-1
Sequence 1, Application US/09370102
Patent No. 6265547
GENERAL INFORMATION:
APPLICANT: Yue, Henry

APPLICANT: Au-Young, Janice
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
FILE REFERENCE: PF-0599 US
CURRENT APPLICATION NUMBER: US/09/370,102
CURRENT FILING DATE: 1999-08-06
EARLIER APPLICATION NUMBER: 09/151,611
EARLIER FILING DATE: 1998-09-11
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 233
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 1974337
US-09-370-102-1

Query Match 11.3%; Score 116; DB 4; Length 233;
Best Local Similarity 32.4%; Pred. No. 2.7e-05;
Matches 36; Conservative 24; Mismatches 41; Indels 10; Gaps 5;

QY 13 ATLNSNTATKG---RYYLEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLSS 68
DB 92 ATVAAFASSEGHSHRPVELPKTDEG---LGFNVGKGEONSPYISRIIPGVADR-HG 147

QY 69 KIQAGDEVHINEVTLSSRKE-AVSLVKGSYKTLRLVYRSLSPPVTSLE 118
DB 148 GLKRGDQLSLVNGSVGEHHEKAVELLKAKADSVKLVVR-VTPKYLEEME 197

RESULT 12
US-09-151-611-3
Sequence 3, Application US/09151611
Patent No. 5958731
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Au-Young, Janice
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
FILE REFERENCE: PF-0599 US
CURRENT APPLICATION NUMBER: US/09/151,611
CURRENT FILING DATE: 1998-09-11
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PERL Program
SEQ ID NO 3
LENGTH: 297
TYPE: PRT
ORGANISM: Caenorhabditis elegans
FEATURE:
OTHER INFORMATION: 91685067
US-09-151-611-3

Query Match 10.7%; Score 110; DB 2; Length 297;
Best Local Similarity 29.5%; Pred. No. 0.00019;
Matches 36; Conservative 28; Mismatches 44; Indels 14; Gaps 6;

QY 13 ATLNSNTATKG---RYYLEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLSS 68
DB 164 ATVAAFASSEGHSHRPVELPKTDEG---LGFNVGKGEONSPYISRIIPGVADR-HG 239
QY 69 KIQAGDEVHINEVTLSSRKEAVSLVKGSYKTLRLVYRSLSPPVTSLE--FPQHPOR 126
DB 240 GLKRGDQLSLVNGSVGEHHEKAVDLKSAVGSVKLVIRYM-PKILDEMERFE---ROR 295

QY 127 MP 128
DB 296 IP 297

RESULT 13

```

US-09-370-102-3
; Sequence 3, Application US/09370102
; Patent No. 6265547
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Au-Young, Janice
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
; FILE REFERENCE: PF-0599 US
; CURRENT APPLICATION NUMBER: US/09/370,102
; CURRENT FILING DATE: 1999-08-06
; EARLIER APPLICATION NUMBER: 09/151,611
; EARLIER FILING DATE: 1998-09-11
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE: -
; OTHER INFORMATION: g1685067
US-09-370-102-3

Query Match      10.7%; Score 110; DB 4; Length 297;
Best Local Similarity 29.5%; Pred. No. 0.00019;
Matches 36; Conservative 28; Mismatches 44; Indels 14; Gaps 6;

QY   13 ATLNSTNATKG----RYILEAFLEGAGPWGFTLKGLGEHGEPLLIISKVEEGKADTLSS 68
     ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db   184 ATAAFAAABGAHPRIVELPKTDQG---LGFNVMGGKEONSPIYSRIIPGVADR-HG 239

QY   69 KLOAGDEVVHINEVTLSRRKEAVSLVKSGYKTLRLVVRSLSPPVTVSLE--PDPQHQR 126
     ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db   240 GLRKGDLIAVNGVEACHEKAVDLLKSAGSVKVLIRYM-PKLDEMERFE---ROR 295

QY   127 MP 128
     :|
Db   296 IP 297

RESULT 14
US-08-123-161A-14
; Sequence 14, Application US/08123161A
; Patent No. 5449616
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Roberts, Steven L.
; APPLICANT: Anderson, Richard D.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,161A
; FILING DATE: 16-SEP-93
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/946,234
; FILING DATE: 14-SEP-92

US-09-370-102-3
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRF89-11A4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-123-161A-14

Query Match      10.5%; Score 108; DB 1; Length 505;
Best Local Similarity 26.3%; Pred. No. 0.00069;
Matches 47; Conservative 23; Mismatches 45; Indels 64; Gaps

QY   39 GFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEVTLS--SRKEAYSLVKG 97
     |::|||::|::|||::|::||::|::||::|::|||::|::|||::|::|||::|::||
Db   99 GISIKGGRENKMPILISIKFKGLAADQ--TEALFVGDAILLSVNGEDLSSATHDEAVQALKK 157

QY   98 SYKTIRLVV---RSLSP-----PVTVSLEFPDQP-QRMP-----PRTT----- 133
     |::|||::|::|||::|::||::|::||::|::|||::|::|||::|::|||::|::||
Db   158 TCEVTVLEVVKYNKYSPFYFKNSAGGTSGVGDSPASPQLQRQSPGPOTRNLSAKHVPL 217

QY   134 -----SFVSSTADG-----RHWSRCRP-----PWVK 154
     :|||::|::|||::|::|||::|::|||::|::|||::|::|||::|::|||::|::||
Db   218 KMAYVSRRCTPSDPHRYLEICSDAQDTIFLRARDEASARSWAGAIOAINALLFPWK 276

RESULT 15
US-08-483-278-14
; Sequence 14, Application US/08483278
; Patent No. 5686073
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Ervasti, James M.
; APPLICANT: Levellie, Cynthia J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,278
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,161
; FILING DATE: 16-SEP-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRF89-11A5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```

```

;
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-278-14

```

```

Query Match          10.5%; Score 108; DB 1; Length 505;
Best Local Similarity 26.3%; Pred. No. 0.00069;
Matches 47; Conservative 23; Mismatches 45; Indels 64; Gaps 9;

```

```

QY 39 GFTLKGGLGHEGEPILISKVEBGGRADTLSSKLAGDEVYHINEVTLS-SRKEAVSLYKG 97
   |::|||::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 99 GISIKGGRKMKPILISKIFKGLADQ-TEALFYGDAILSYNGEDLSSATHDEAVQALKK 157
   |::|||::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 98 SYKTLRLVY---RSLSP-----PVTVSLEFDPQHP-QRMP---PRTPT----- 133
   |::|||::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 158 TGRKVVLEVKYKKEVSPYFKNSAGCTSVGWDSPPASPLQRPSSPGPQTRNLSEAKHVPL 217
   |::|||::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 134 -----SFSVSTADG-----RHENSCR-----PWYK 154
   |::|||::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 218 KMAVVSRRCPSPDEPHRYLEICASADGGDTIFLRAKDEASRSWAGAIQAIINALLPWYK 276
   |::|||::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Search completed: January 31, 2002, 20:57:10
 Job time: 53 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:59:22 ; Search time 243.53 Seconds
(without alignments)
216.626 Million cell updates/sec

Title: US-09-641-831-2
Perfect score: 1029
Sequence: 1 MRRTEDEHFKPSATLNSNTA.....CIYPTCHNCRNFRAYLSR 190

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 3148936 seqs, 277657034 residues
Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA_Main: *
1: /cgn2_6/ptodata/2/paa/PCrus_COMB.pep: *
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep: *
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep: *
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep: *
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep: *
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep: *
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep: *
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep: *
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep: *
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep: *
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep: *
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep: *
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep: *
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep: *
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep: *
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep: *
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep: *
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep: *
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep: *
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep: *
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep: *
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep: *
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep: *
24: /cgn2_6/ptodata/2/paa/US60_COMB.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	290	28.2	1616	1 PCT-US01-04098A-1172	Sequence 1172, Ap
2	290	28.2	1646	1 PCT-US01-04098A-3140	Sequence 3140, Ap
3	238.5	23.2	189	21 US-09-758-472-9626	Sequence 9626, Ap
4	134	13.0	218	1 PCT-US01-01312-1236	Sequence 1236, Ap
5	134	13.0	218	21 US-09-764-868-1236	Sequence 1236, Ap
6	133	12.9	834	24 US-60-212-413-260	Sequence 260, App
7	133	12.9	1032	24 US-60-258-275-261	Sequence 261, App
8	130.5	12.7	493	24 US-60-173-464-10230	Sequence 10230, A
9	130.5	12.7	493	24 US-60-191-637-12599	Sequence 12599, A

10	130.5	12.7	493	24	US-60-191-681-9887	Sequence 9887, Ap
11	130.5	12.7	734	1	PCT-US01-14827-13022	Sequence 13022, A
12	130.5	12.7	944	24	US-60-167-217-12593	Sequence 12593, A
13	129.5	12.6	750	24	US-60-258-251-58	Sequence 58, Appl
14	129.5	12.6	981	24	US-60-248-505-946	Sequence 946, Appl
15	128.5	12.5	163	1	PCT-US01-01312-964	Sequence 964, App
16	128.5	12.5	163	21	US-09-764-868-964	Sequence 964, App
17	128.5	12.5	207	1	PCT-US00-32990-19	Sequence 19, Appl
18	128.5	12.5	207	24	US-60-172-354-19	Sequence 19, Appl
19	128.5	12.5	399	24	US-60-173-664-23192	Sequence 23192, A
20	128.5	12.5	400	24	US-60-191-637-29251	Sequence 29251, A
21	128.5	12.5	400	24	US-60-191-681-23602	Sequence 23602, A
22	127	12.3	327	16	US-09-206-645-10	Sequence 10, Appl
23	126.5	12.3	111	12	US-08-805-819-23	Sequence 17, Appl
24	126.5	12.3	406	14	US-09-028-724-17	Sequence 159, App
25	126.5	12.3	406	19	US-09-502-967-159	Sequence 601, App
26	126	12.2	195	24	US-60-167-245-601	Sequence 14024, A
27	126	12.2	195	24	US-60-173-664-14024	Sequence 17123, A
28	126	12.2	195	24	US-60-191-637-17123	Sequence 13553, A
29	126	12.2	195	24	US-60-191-681-13553	Sequence 33346, A
30	126	12.2	204	16	US-09-270-767-33346	Sequence 48563, A
31	126	12.2	204	16	US-09-270-767-48563	Sequence 191322, A
32	126	12.2	204	16	US-09-270-849B-191322	Sequence 17041, A
33	125.5	12.2	129	22	US-09-834-366-17041	Sequence 17041, A
34	125.5	12.2	136	22	US-09-834-366-14514	Sequence 14514, A
35	125.5	12.2	136	24	US-60-197-873-15955	Sequence 15955, A
36	125.5	12.2	136	24	US-60-197-873-15955	Sequence 15955, A
37	125.5	12.2	139	22	US-09-834-366-15955	Sequence 3059, Appl
38	125.5	12.2	139	24	US-60-197-873-15955	Sequence 40, Appl
39	125.5	12.2	329	16	US-09-206-645-3	Sequence 34, Appl
40	125.5	12.2	376	1	PCT-US01-18569-1059	Sequence 10, Appl
41	124.5	12.1	153	1	PCT-US00-11664-40	
42	124.5	12.1	223	1	PCT-US00-11664-34	
43	124.5	12.1	223	21	US-09-721-975-34	
44	124.5	12.1	457	1	PCT-US00-11664-10	
45	124.5	12.1	457	21	US-09-721-975-10	

ALIGNMENTS

RESULT 1
PCT-US01-04098A-1172
Sequence 1172, Application PC/TUS0104098A
GENERAL INFORMATION:
APPLICANT: Hysq, Inc.
FILE OF INVENTION: Novel Nucleic Acids and Polypeptides
TITLE REFERENCE: 21272-029
CURRENT APPLICATION NUMBER: PCT/US01/04098A
CURRENT FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: Not Yet Assigned
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 09/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 3960
SOFTWARE: Custom
SEQ ID NO 1172
LENGTH: 1616
TYPE: PRT

```
; ORGANISM: Homo sapiens
PCT-US01-04098A-1172

Query Match      28.2%; Score 290; DB 1; Length 1616;
Best Local Similarity 49.6%; Pred. No. 6e-19;
Matches 70; Conservative 18; Mismatches 43; Indels 10; Gaps 5;

QY 13 ATLSNTATKGRYILEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLLSSKLOA 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 14 AEAETRAADGRLV--EVLSGAPWGFTLKGREHGEPLVITKEEGSKAAV-DKLLA 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 73 GDEVHINEVTLSSSKRAVSLVKGYSKTLRLVRSLS-----PPVTVSLEFDPQHPQ-R 126
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71 GDEIVGINDIGLSGFQEAICLVKGSHTKLKLVKRRLSGWRHPSHWATKFSDSHPELA 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 127 MPPTRTFSVSVDGRHEWS 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 131 ASPFTSTGCPSPWS-GRHHAS 150
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
PCT-US01-04098A-3140
; Sequence 3140, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 3140
; LENGTH: 1646
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-3140

Query Match      28.2%; Score 290; DB 1; Length 1646;
Best Local Similarity 49.6%; Pred. No. 6.2e-19;
Matches 70; Conservative 18; Mismatches 43; Indels 10; Gaps 5;

QY 13 ATLSNTATKGRYILEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLLSSKLOA 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 44 AEAETRAADGRLV--EVLSGAPWGFTLKGREHGEPLVITKEEGSKAAV-DKLLA 100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 73 GDEVHINEVTLSSSKRAVSLVKGYSKTLRLVRSLS-----PPVTVSLEFDPQHPQ-R 126
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 101 GDEIVGINDIGLSGFQEAICLVKGSHTKLKLVKRRLSGWRHPSHWATKFSDSHPELA 160
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 127 MPPTRTFSVSVDGRHEWS 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161 ASPFTSTGCPSPWS-GRHHAS 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Query Match      23.2%; Score 238.5; DB 21; Length 189;
Best Local Similarity 57.5%; Pred. No. 4.6e-15;
Matches 46; Conservative 16; Mismatches 17; Indels 1; Gaps 1;

QY 35 GAPWGTFLKGGLEHGEPLIISKVEEGKADTLLSSKLOAGDEVVHINEVTLSSSKRAVSL 94
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 19 GGTWGTFLKGGLEHGEPLVSKIEDGGA-ALSOQKRTGDELVNINGTPLYGSRQEAAIL 77
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 95 VKGSYKTLRLVRSLSPPVT 114
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78 IKGSFRILKLVRRNAPVS 97
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
PCT-US01-01312-1236
; Sequence 1236, Application PC/TUS0101312
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01312
; CURRENT FILING DATE: 2001-05-09
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1236
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (134)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (137)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (164)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
PCT-US01-01312-1236

Query Match      13.0%; Score 134; DB 1; Length 218;
```



```
; Sequence 10230, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10230
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Drosophila
US-60-173-464-10230

Query Match      12.7%; Score 130.5; DB 24; Length 493;
Best Local Similarity 34.0%; Pred. No. 0.00093;
Matches 36; Conservative 16; Mismatches 39; Indels 15; Gaps 3;

QY 33 EGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLGAGDEVVHINEVTLS--RKEA 91
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 80 ERGSTYGTVRGREGHTGFFVSHVEHGEAHL--KGLRIGDQILRINGFRLLDDAVHKEF 137
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||

QY 92 VSLVKGSKYKTLRLVRSLSPPVTVSLEFDPQHPQMPPTRTTSFSV 137
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 138 IQLVAGQDR-----VTLKRVGVGMPLPVRDLPEERLSWSV 171

RESULT 9
US-60-191-637-12599
; Sequence 12599, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12599
; LENGTH: 493
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-637-12599

Query Match      12.7%; Score 130.5; DB 24; Length 493;
Best Local Similarity 34.0%; Pred. No. 0.00093;
Matches 36; Conservative 16; Mismatches 39; Indels 15; Gaps 3;

QY 33 EGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLGAGDEVVHINEVTLS--RKEA 91
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 80 ERGSTYGTVRGREGHTGFFVSHVEHGEAHL--KGLRIGDQILRINGFRLLDDAVHKEF 137
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||

QY 92 VSLVKGSKYKTLRLVRSLSPPVTVSLEFDPQHPQMPPTRTTSFSV 137
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 138 IQLVAGQDR-----VTLKRVGVGMPLPVRDLPEERLSWSV 171

RESULT 10
US-60-191-681-9887
; Sequence 9887, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
; FILE REFERENCE: cl000390
```

```
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9887
; LENGTH: 493
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-681-9887

Query Match      12.7%; Score 130.5; DB 24; Length 493;
Best Local Similarity 34.0%; Pred. No. 0.00093;
Matches 36; Conservative 16; Mismatches 39; Indels 15; Gaps 3;

QY 33 EGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLGAGDEVVHINEVTLS--RKEA 91
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 80 ERGSTYGTVRGREGHTGFFVSHVEHGEAHL--KGLRIGDQILRINGFRLLDDAVHKEF 137
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||

QY 92 VSLVKGSKYKTLRLVRSLSPPVTVSLEFDPQHPQMPPTRTTSFSV 137
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 138 IQLVAGQDR-----VTLKRVGVGMPLPVRDLPEERLSWSV 171

RESULT 11
PCT-US01-14827-13022
; Sequence 13022, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hysed, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 13022
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (636)..(651)
; OTHER INFORMATION: LIM domain proteins domain identified by eMatrix, accession
; OTHER INFORMATION: number BL00478B, p-value=4.250e-12, raw score of 14.79
; NAME/KEY: DOMAIN
; LOCATION: (558)..(734)
; OTHER INFORMATION: LIM domain containing proteins domain identified by Pfam,
; OTHER INFORMATION: accession name LIM, E-value=1.4e-28, Pfam score of 106.0
PCT-US01-14827-13022

Query Match      12.7%; Score 130.5; DB 1; Length 734;
Best Local Similarity 30.2%; Pred. No. 0.0016;
Matches 42; Conservative 20; Mismatches 68; Indels 9; Gaps 3;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLGAGDEVVHINEV-TLSSSRKE 90
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 14 LTGPGWGFRLGGKDFNMPLTISRITPGSKA--AQSQLSQGLVVAIDGVNDTMTHLE 71
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||

QY 91 AVSLVKGSKYKTLRLVRSLSPPVTVSLEFDPQH-----PQMPPTRTTSFSVSADGRH 144
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 72 AQNKIKASYNLSLTQKSKRPIPTTAPPVOTPLPVPHPQDPAIDNGSLVAPSPSP 131
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||

QY 145 EMSCRPPVYKWSPRETWA 163
| | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||: |||
Db 132 EARASPGTPTGTPELRPTFS 150

RESULT 12
US-60-167-217-12593
; Sequence 12593, Application US/60167217
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:58:12 ; Search time 40.57 Seconds

(Without alignments)
356.745 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029

Sequence: 1 MMRTEDEPKPSATLNSMTA.....CIPTQHNWCRNFKRAYLSR 190

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:**

1: p1r1:**

2: p1r2:**

3: p1r3:**

4: p1r4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	290	28.2	1616	2 137183	gene APXL protein
2	128.5	12.5	852	2 T10811	channel associated
3	128.5	12.5	870	2 G01974	channel associated
4	128	12.4	455	2 A55050	enigma - human
5	127	12.3	327	2 JCA385	LIM protein - rat
6	122.5	11.9	913	1 A41109	protein-tyrosine-p
7	121.5	11.8	330	2 S71828	gene r11 protein -
8	121	11.7	911	2 I56552	synapse-associated
9	120.5	11.7	1171	2 T42372	probable guanylate
10	118.5	11.5	1256	2 JEO209	brain-specific ang
11	117.5	11.4	517	2 T47134	hypothetical prote
12	116	11.3	628	2 T09458	numb-binding prote
13	116	11.3	728	2 T09457	numb-binding prote
14	115	11.2	431	2 T16191	hypothetical prote
15	113.5	11.0	904	2 I38757	homolog of Drosoph
16	113.5	11.0	926	2 I38756	homolog of Drosoph
17	113.5	11.0	1012	2 T23160	hypothetical prote
18	110.5	10.7	317	2 T27179	hypothetical prote
19	109	10.6	624	2 T19650	hypothetical prote
20	109	10.6	1026	2 T19631	hypothetical prote
21	109	10.6	1131	2 T15617	hypothetical prote
22	108.5	10.5	322	2 S19649	cysteine proteinas
23	108.5	10.5	720	2 A45436	synapse-associated
24	108	10.5	505	2 A53214	dystrophin-associ
25	107	10.4	724	2 JH0800	postsynaptic densi
26	107	10.4	767	2 T09599	postsynaptic densi
27	106.5	10.3	1112	2 T32733	AMPA glutamate rec
28	105	10.2	926	1 A41105	protein-tyrosine-p
29	104.5	10.2	503	2 I84771	syntrophin-1 - mou

30	103.5	10.1	538	2 I59291	beta1-syntrophin -
31	103.5	10.1	960	1 A39651	discs-large tumor
32	103.5	10.1	1277	2 T14152	synaptic scaffold
33	103	10.0	505	2 S62894	alpha-syntrophin -
34	102.5	10.0	2450	2 S71625	protein-tyrosine-p
35	99.5	9.7	1387	2 JCS502	G-protein signalin
36	98.5	9.6	1367	2 T13703	tama protein - fru
37	97.5	9.5	423	2 T21570	hypothetical prote
38	97.5	9.5	440	2 T21568	hypothetical prote
39	96.5	9.4	488	2 I51379	syntrophin - Pacif
40	95	9.2	723	2 T14765	hypothetical prote
41	92	8.9	390	2 T28036	hypothetical prote
42	90	8.7	413	2 S77339	hypothetical prote
43	90	8.7	2294	2 I67630	protein tyrosine p
44	90	8.7	2466	2 I67629	protein tyrosine p
45	88	8.6	2490	1 A54971	protein-tyrosine-p

ALIGNMENTS

RESULT 1
137183
gene APXL protein - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence-revision 29-May-1998 #text-change 21-Jul-2000
C:Accession: 137183
R:Schlaefli, M.V.; Bassi, M.T.; Rugerli, E.I.; Rennerli, A.; Galli, L.; Ballabio, A.
Hum. Mol. Genet. 4, 373-382, 1995
A:Title: Cloning of a human homologue of the Xenopus laevis APX gene from the ocular
A:Reference number: 137183; MUID:5315933
A:Accession: 137183
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1616 <RES>
A:Cross-references: EMBL:X83543; NID:9790999; PIDD:CA58534.1; PID:g1181628
C:Genetics:
A:Gene: GDB:APXL
A:Cross-references: GDB:582527
A:Map position: XP22.3-XP22.3
C:Superfamily: GLGF domain homology
F:32-104/Domain: GLGF domain homology <GLG>

Query Match 28.2%; Score 290; DB 2; Length 1616;
Best Local Similarity 49.6%; Pred. No. 4.2e-19;
Matches 70; Conservative 18; Mismatches 43; Indels 10; Gaps 5;

QY 13 ATLNSNTKRGRIYILEFLEGAPMGFTLKGLEHGPILLISKEVEEGKADTLSSKIQ 72
DB 14 AEAEFRADGGRLV--EVLSSGAPMGFTLKGREHGPVITKIEEGSKAAV-DKLLA 70
QY 73 GDEVHINEVTLSSSRKDAVSLVKSQYTLRLVYRSL-----PPVTVSLFDPQHPQ-R 126
DB 71 GDEIVGINDIGLSGRDAICLVKSHKTLKLVYKRREBELGMRPMSWATKFSQHPBLA 130
QY 127 MPPRTISFSVSTADGRHWS 147
DB 131 ASPFTSGCPSWS-GRHWS 150

RESULT 2
T10811
channel associated protein of synapse 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence-revision 16-Jul-1999 #text-change 21-Jan-2000
C:Accession: T10811
R:Itte, M.; Hata, Y.; Takai, Y.
Submitted to the EMBL Data Library, April 1996
A:Description: Cloning of new isoforms of PSD-95/SAP90 related genes.
A:Reference number: 217166
A:Accession: T10811
A:Status: preliminary; translated from GB/EMBL/DBJ

Qy 32 LEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLQAGDEWIIIN-EVTLSSRKE 90
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 LOGCPWFGRVLGGKDFEQPLAISRVTPGSKA -AIANLCIGDLITADGEDTSSMTHLE 65

Oy 91 AVSLVKGSYKTLRLVV-----RSLSPPTVLSLEFPDHPQRN-----PPR 130

Db 66 AONKIGCDNNMTLTVSRSEOKTMSPLVT---EGRKHPKKNLASEPOEVLHIGSAHNR 122
 QY 131 TRTSFVSSTADG 142
 Db 123 SAMFTASPPAG 134

RESULT 6
 A41109
 protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 3 - human
 N:Alternate names: PTPH1
 C:Species: Homo sapiens (man)
 C>Date: 27-Mar-1992 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
 C:Accession: A41109; 155698
 R:Yang, Q.; Tonks, N.K.
 A:Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with hc
 A:Reference number: A41109; MUID:91296738
 A:Accession: A41109
 A:Molecule type: mRNA
 A:Residues: 1-913 <YAN>
 A:Cross-references: GB:M64572; NID:q179912; PIDN:AAA35647.1; PID:q179913
 R:Kuta, S.; Iton, F.; Hinoda, Y.; Toyota, M.; Makiyuchi, Y.; Imai, K.; Yachi, A.
 J. Galectin-1, 29, 727-732, 1994
 A:Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA
 A:Reference number: 155698; MUID:95179278
 A:Accession: 155698
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 899-913 <RSS>
 A:Cross-references: GB:S76309; NID:q913165; PIDN:AAB33583.1; PID:q913166
 C:Genetics:
 A:Gene: GDB:PTPN3
 A:Cross-references: GDB:131386; OMIM:176877
 A:Map position: 9q31-9q31
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology; F
 C:Keywords: phosphoprotein, phosphoric monoester hydrolase; tyrosine-specific phosphatase
 F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
 F:516-890/Domain: GLGF domain homology <GLG>
 F:670-890/Domain: protein-tyrosine-phosphatase homology <PTP>
 F:842/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:848/Binding site: substrate phosphate (Arg) #status predicted

Query Match 11.9%; Score 122.5; DB 1; Length 913;
 Best Local Similarity 27.7%; Pred. No. 0.002; Mismatches 44; Indels 19; Gaps 4;
 Matches 33; Conservative 23; Mismatches 44; Indels 19; Gaps 4;

QY 6 EDFRPSATLNSNTATKGRY-----IYLEAFLEGGAPWFTLKGLEHGEPLI 53
 Db 479 DDFR--VTKGSTEDASQYCKDNKNDSTLVLRITPDDGKFGFLKGGVQDKMPLV 536

QY 54 ISKVEGKADTLSSKIQAGDEVVHINEVTL--SKREAVSLVSGSYKT----LRLVLR 107
 Db 537 VSRINPSPADTCIPKINEGDOIVLINGRDISEHTHQDVVFIRASRSHRELALVIR 595

RESULT 7
 S71828
 gene r11 protein - rat
 N:Alternate names: LIM-domain protein r11
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 05-Nov-1999
 C:Accession: S71828; S39246
 R:Schaefer, R.
 submitted to the EMBL Data Library, March 1996
 A:Reference number: S71828
 A:Accession: S71828
 A:Molecule type: mRNA
 A:Residues: 1-330 <SCH>
 A:Cross-references: EMBL:X76454; NID:q755780; PIDN:CAA53992.1; PID:q887580
 C:Genetics:

A:Gene: r11
 C:Superfamily: LIM metal-binding repeat homology; GLGF domain homology
 F:8-80/Domain: GLGF domain homology <GLG>
 F:255-305/Domain: LIM metal-binding repeat homology <LIM2>

Query Match 11.8%; Score 121.5; DB 2; Length 330;
 Best Local Similarity 32.0%; Pred. No. 0.00065;
 Matches 41; Conservative 15; Mismatches 51; Indels 21; Gaps 5;

QY 32 LFGGAPWFTLKGLEHGEPLIISKVEGKADTLSSKIQAGDEVVHIN-EVTLSSSRE 90
 Db 7 LRGPSPWGFVLGGRFNSAPLTSRVHAGSKA--ALALCPGDSIQAINGESTELMTILE 64

QY 91 AVSLVKSYSKTLRLVVRSLSPYTVS-----LEFPPQHPKMPRTKTS--F 135
 Db 65 AONRIGCHDHLTL---SVSRPENKNWSPSPNDKAQHRHIDPEADQSPATSRSSIS 121

QY 136 SVSTADGR 143
 Db 122 GISLEDR 129

RESULT 8
 156552
 synapse-associated protein 97 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jan-2000
 C:Accession: 156552
 R:Miller, B.M.; Kistner, U.; Voh, R.W.; Cases-Langhoff, C.; Becker, B.; Gundelfinger, J.
 J. Neurosci. 15, 2354-2366, 1995
 A:Title: Molecular characterization and spatial distribution of SAP97, a novel presyn
 A:Reference number: 156552; MUID:95198112
 A:Accession: 156552
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-911 <RES>
 A:Cross-references: EMBL:U14950; NID:g642455; PIDN:AAA79976.1; PID:g642456
 C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
 F:229-307/Domain: GLGF domain homology <GLG1>
 F:323-401/Domain: GLGF domain homology <GLG2>
 F:587-645/Domain: SH3 homology <SH3>
 F:722-899/Domain: guanylate kinase homology <GKI>

Query Match 11.8%; Score 121; DB 2; Length 911;
 Best Local Similarity 25.8%; Pred. No. 0.0028;
 Matches 42; Conservative 30; Mismatches 55; Indels 36; Gaps 7;

QY 3 RTEDFHKPSATLNSNTATKGRYTL-EAFL-----EGGAPWFTLKGFL 46
 Db 424 QSVNHNWSPSSYLQOTPASPARYSPIKAVLGDDEITREPRLVLRHSGTGLGNIVG- 482

QY 47 EHGEPLIISKVEGKADTLSSKIQAGDEVVHINEVTL--SSKREAVSLVSGSYKTRLV 105
 Db 483 EDGECIFSLFLLAGPAD-LSGELPKGDRIISVSVLDRAASHQMAALKNK----- 534

QY 106 VRSLSPPVTVSLERDPQHPQMPPR-----TRTSFVSSTADG 142
 Db 535 ---GQAVTIVAYRPEYSRFEAKIHDLRETMNNSVSSGSG 573

RESULT 9
 T42372
 probable guanylate kinase (EC 2.7.4.8) 1, membrane-associated, splice form b - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
 C:Accession: T42372
 R:Dobrosotskaya, I.; Guy, R.K.; James, G.L.
 J. Biol. Chem. 272, 31589-31597, 1997
 A:Title: MAGI-1: a membrane-associated guanylate kinase with a unique arrangement of
 A:Reference number: 222139; MUID:98058950
 A:Accession: T42372

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1171 <DOB>
A:Cross-references: EMBL:AF027503; NID:g2702346; PID:g2702347; PIDN:AAB91995.1
A:Experimental source: strain C57 Black/6 x CBA
C:Genetics:
A:Gene: Magi-1
C:Superfamily: WW repeat homology
C:Keywords: alternative splicing; phosphotransferase
F:300-337/Domain: WW repeat homology <WR1>
F:347-384/Domain: WW repeat homology <WR2>

Query Match 11.7%; Score 120.5; DB 2; Length 1171;
Best Local Similarity 26.5%; Pred. No. 0.0044;
Matches 35; Conservative 22; Mismatches 46; Indels 29; Gaps 5;
QY 9 HKPS--ATLNSNTATKGRY-----IYLEAFLEGAPWGFTLKGGLEHGE 50
DB 998 HAPSOQGTQETRTTKPKQDSOFKFGPQAQEQDYTVLELGRGAKGFGFSLRGGRYNNM 1057
QY 51 PLIISKVEGGKADTLSSKLAQGDVHHN-EVTLSSSRKEAVSLKGYKTLRLVVR-- 107
DB 1058 DLYVRLAEDGPAER-CGKMRIIGDEILGETTKNMKHSRAIELIKNGRRVRLFLRG 1116
QY 108 -----SLSP 112
DB 1117 DGSVPEYAMIPP 1128

RESULT 10
JEO209
brain-specific angiogenesis inhibitor-associated protein 1 - human
N:Alternate names: BAI1-associated protein 1; BAP1 [misnomer]
C:Species: Homo sapiens (man)
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 18-Aug-2000
C:Accession: JEO209
R:Shiratsuchi, T.; Futamura, M.; Oda, K.; Nishimori, H.; Nakamura, Y.; Tokino, T.
Biochem. Biophys. Res. Commun. 247, 597-604, 1998
A:Title: Cloning and characterization of BAI1-associated protein 1: A PDZ domain-containing
A:Reference number: JEO209; MUID:98321173
A:Accession: JEO209
A:Molecule type: mRNA
A:Residues: 1-1256 <SHI>
A:Cross-references: GB:AB010894; NID:g3370997
A:Experimental source: brain
C:Genetics:
A:Gene: GDB:BAIAP1; BAP1
A:Cross-references: GDB:9864783
A:Map position: 3p14.1-3p14.1
C:Superfamily: WW repeat homology
F:300-337/Domain: WW repeat homology <WR1>
F:359-396/Domain: WW repeat homology <WR2>

Query Match 11.5%; Score 118.5; DB 2; Length 1256;
Best Local Similarity 30.4%; Pred. No. 0.0075;
Matches 31; Conservative 22; Mismatches 40; Indels 9; Gaps 4;
QY 7 DFHKPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGGLEHGEPLIISKVEEGKADTL 66
DB 1138 EFKAQATQEQDYF-----VELE---RCAGKGFGFSLRGGRYNNMDLYVRLAEDGPAER- 1189
QY 67 SSKLAQGDVHHN-EVTLSSSRKEAVSLKGYKTLRLVVR 107
DB 1190 CGKMRIIGDEILGETTKNMKHSRAIELIKNGRRVRLFLK 1231

RESULT 11
T47134
hypotheical protein DKF2p761i2312.1 - human
C:Species: Homo sapiens (man)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47134
R:Ansoorge, W.; Wirkner, U.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24375
A:Accession: T47134
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-517 <AAA>
A:Cross-references: EMBL:AL161971
A:Experimental source: adult amygdala; clone DKF2p761i2312
C:Genetics:
A:Note: DKF2p761i2312.1

Query Match 11.4%; Score 117.5; DB 2; Length 517;
Best Local Similarity 23.4%; Pred. No. 0.0028;
Matches 46; Conservative 31; Mismatches 65; Indels 55; Gaps 10;
QY 6 EDFHKPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGGLEHGEPLIISKVEEGKADT 65
DB 49 EPFYSGERTVIRRTVGGF-----GLSIKGAHNPVYVYSKISKEQRAE- 94
QY 66 LSSKLAQGDVHHN-EVTLSSSRKEAVSLKGYKTLRLVVRSLPVPVTVSLEFDPQHPQ 125
DB 95 LSGLLFIGDALIQINGINVKCRHEEV-----QVLRNAGEEVTIVSFLKRAPA 144
QY 126 --RMP-----PRTRTSFSVSTA--DGRH-----EWSCRPPVWVWWSRPTWAA 164
DB 145 FLKLPLNEDCACAPSDQSSGTSPLCDGSLHNPNTDITLSCSS-----WPTSP--GL 197
QY 165 RWPQKGC---IYPTQHN 178
DB 198 RWEKRWCDLRLIPLLHS 214

RESULT 12
T09458
numb-binding protein LNXp70 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09458
R:Dho, S.E.; Jacob, S.; Wolting, C.D.; French, M.B.; Rohrschneider, L.R.; McGlade, C.
J. Biol. Chem. 273, 9179-9187, 1998
A:Title: The mammalian Numb phosphotyrosine binding domain: Characterization of bindi
A:Reference number: Z16678; MUID:98204916
A:Accession: T09458
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-628 <DHO>
A:Cross-references: EMBL:AF034746; NID:g3041880; PIDN:AAC40076.1; PID:g3041881
A:Experimental source: Brain
C:Genetics:
A:Gene: lnx

Query Match 11.3%; Score 116; DB 2; Length 628;
Best Local Similarity 25.0%; Pred. No. 0.0051;
Matches 39; Conservative 27; Mismatches 52; Indels 38; Gaps 8;
QY 10 KPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGGLEHGE---PLIISKVEEGKADTL 66
DB 399 KPAATCHEKV-----VSVWKDPSESIGMTVGGGASHREWDLPYIVISVEPGG-VISR 449
QY 67 SSKLAQGDVHHN-EVTLSS--SRKEAVSLKGYKTLRLVVRSLPVPVTVSLEFDPQHPQ 125
DB 450 DGRIKTGILLNNGIELTEVSRTEAVAILKSAPSV-----VLKALEVKEQEAQ 499
QY 126 RMPTRTSFSVSTADGRH-----EWSCRPPVWVW 155
DB 500 E-----DCSPAALDSNNHNPVTPGDWS--PSWVNM 526
RESULT 13

T09457
 numb-binding protein LNXp80 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Sep-2000
 C:Accession: T09457
 R:Bio. S.E.: Jacob, S.; Molling, C.D.; French, M.B.; Rohrrechner, L.R.; McGlade, C.J.
 J. Biol. Chem. 273, 9179-9187, 1998
 A:Title: The mammalian Numb phosphotyrosine binding domain: Characterization of binding
 A:Reference number: 216678; MUID:98204916
 A:Accession: T09457
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-728 <DHO>
 A:Cross-references: EMBL:AF034745; NID:g3041878; PIDN:AMC40075.1; PID:g3041879
 A:Experimental source: developmental stage: embryo
 C:Genetics:
 A:Gene: Lnx
 C:Superfamily: RING finger homology
 F:41-88/Domain: RING finger homology <RRN>

Query Match 11.3%; Score 116; DB 2; Length 728;
 Best Local Similarity 25.0%; Pred. No. 0.0062;
 Matches 39; Conservative 27; Mismatches 52; Indels 38; Gaps 8;
 QY 10 KPSATLNSMTATKGRITYEAFLGSGAPWGFTLKGLEHGE--PLISKVEEGKADTL 66
 DB 499 KPAATCHEVY-----VSWKDPSESIGMTVGGASHREMDPLIVISVEPGG-VISR 549
 QY 67 SSKLAGDEVHINEVTLSS-SRKEAVSLVKGSKYKTLRLVRSLSPPVVSLEFDPQHPQ 125
 DB 550 DSRKTKGDLILNNGELTEVSRTEAVILKSPSSV-----VLALEVKQENQ 599
 QY 126 RMPPTRTISFSVSTADGRH-----EMSCRPPVWK 155
 DB 600 E-----DCSPALDLSNHNVTTPGDM--PSWVM 626

RESULT 14
 T16191
 hypothetical protein F27D9.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 A:Accession: T16191
 R:Bentley, D. submitted to the EMBL Data Library, February 1996
 A:Description: The sequence of C. elegans cosmid F27D9.
 A:Reference number: 218473
 A:Accession: T16191
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-431 <BEN>
 A:Cross-references: EMBL:U49629; NID:g1203924; PID:g1203931; PIDN:AAA93388.1; CESP:F27D9
 C:Genetics:
 A:Gene: CESP:F27D9.8
 A:introns: 8/1; 52/3; 74/1; 96/1; 128/2; 149/3; 199/2; 255/3; 297/3; 319/3; 350/3; 390/3

Query Match 11.2%; Score 115; DB 2; Length 431;
 Best Local Similarity 27.2%; Pred. No. 0.0039;
 Matches 43; Conservative 24; Mismatches 51; Indels 40; Gaps 7;
 QY 29 EAFLEGG-----APWGFTLKGLEHGE--PLISKVEEGKADTLSSKLAGDE 75
 DB 27 ETFLELATNOVVIVKRPDSGFGSLIKGSENNAMPVIVISKIPKLPAD-CEGLFTGDA 85
 QY 76 VVHINEVTL-SSSRKEAVSLVKGSKYKTLRLVYR-----SLSPVTVSLEFDP 121
 DB 86 IYEVNIGISTEGSHDEVVKNKSSGDOYTLGVNHTFHTKTPFLKPAOSLOPDGTLDDLFDS 145
 QY 122 Q-----HPQMPPTRTISFSVSTADGRHESCRPPVWK 155
 DB 146 RSTRSHKAQSESRSLSDSKWQOHIDGREK-----KW 175

RESULT 15
 I38757
 homolog of Drosophila discs large protein, isoform 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
 C:Accession: I38757
 R:Jue, R.A.; Marfatta, S.M.; Branton, D.; Chishti, A.H.
 Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
 A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila
 A:Reference number: I38756; MUID:95024052
 A:Accession: I38757
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-904 <RES>
 A:Cross-references: EMBL:U13897; NID:g558437; PIDN:AAA50599.1; PID:g558438
 C:Genetics:
 A:Gene: GDB:DLG1
 A:Cross-references: GDB:393278; OMIM:601014
 A:Map position: 3q29-3q29
 C:Superfamily: discs-large tumor suppressor; GIGF domain homology; guanylate kinase h
 C:Keywords: alternative splicing; duplication
 F:229-307/Domain: GIGF domain homology <GIG1>
 F:324-402/Domain: GIGF domain homology <GIG2>
 F:588-646/Domain: SH3 homology <SH3>
 F:715-892/Domain: guanylate kinase homology <GKI>

Query Match 11.0%; Score 113.5; DB 2; Length 904;
 Best Local Similarity 26.2%; Pred. No. 0.014;
 Matches 42; Conservative 27; Mismatches 54; Indels 37; Gaps 8;
 QY 7 DFN-KPSATLNSMTATKGRITYL-EAFL-----EGGAPWGFTLKGLEHG 49
 DB 428 DNVSPSSFLGQTPASPARSPVSKAVLGDDETTRPRKYVLHKGSTGGLFNIVGG-EDG 486
 QY 50 EPLISKVEEGKADTLSSKLAGDEVHINEVTL-SSSRKEAVSLVKGSKYKTLRLVNS 108
 DB 487 EGFLIFTLAGPAD-LSGELRKGDRTISVNSVDLRAASHQAAALKNK----- 535
 QY 109 LSPVTVSLEFDPQHPQMPPTRT-----TSFSVSTADG 142
 DB 536 -GQAVTVIVAGYRREYSRFEAKIHDLRQNMNNSISSGSG 574

Search completed: January 31, 2002, 20:59:16
 Job time: 64 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:05:37 ; Search time 26.99 Seconds
(without alignments)
258,107 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MMRTEDFHKPSATLSNNTA.....CIYPOHNCRNKRAYLR 190

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	290	28.2	1616 1 APYL_HUMAN	01396 homo sapien
2	134	13.0	330 1 RIL_CHICK	099722 gallus gall
3	128.5	12.5	852 1 DLG2_RAT	063622 rattus norv
4	128.5	12.5	870 1 DLG2_HUMAN	015700 homo sapien
5	127	12.3	327 1 PDL1_RAT	P52944 rattus norv
6	122.5	11.9	913 1 PTN3_HUMAN	P26045 homo sapien
7	121.5	11.8	330 1 RIL_RAT	P36202 rattus norv
8	121	11.8	911 1 DLG1_RAT	P62696 rattus norv
9	119.5	11.6	849 1 DLG3_MOUSE	P70175 mus musculu
10	119.5	11.6	849 1 DLG3_RAT	062935 rattus norv
11	117.5	11.4	329 1 PDL1_HUMAN	000151 homo sapien
12	113.5	11.0	904 1 DLG1_HUMAN	012859 homo sapien
13	112	10.9	817 1 DLG3_HUMAN	Q92796 homo sapien
14	110	10.7	328 1 RIL_HUMAN	P50479 homo sapien
15	109	10.6	326 1 PDL1_MOUSE	070400 mus musculu
16	109	10.6	1026 1 PTN1_CAEEL	P28191 caenorhabd1
17	108.5	10.5	322 1 CYSL_HUMAN	P13277 homarus ame
18	107	10.4	724 1 DLG4_MOUSE	062108 mus musculu
19	107	10.4	724 1 DLG4_RAT	P31016 rattus norv
20	107	10.4	767 1 DLG4_HUMAN	P78352 homo sapien
21	105	10.2	926 1 PTN4_HUMAN	P29074 homo sapien
22	103.5	10.1	960 1 DLG1_DROME	P31007 drosophila
23	99.5	9.7	1387 1 RGSC_RAT	008774 rattus norv
24	92.5	9.0	1447 1 RGSC_HUMAN	014974 homo sapien
25	90.5	8.8	643 1 RHOP_MOUSE	061085 mus musculu
26	90	8.7	2485 1 PTND_HUMAN	Q12923 homo sapien
27	87.5	8.5	323 1 CYSL_HUMAN	P25782 homarus ame
28	87.5	8.5	933 1 ZO3_HUMAN	095049 homo sapien
29	86	8.4	888 1 ZO3_CANFA	062683 canis fami1
30	84	8.2	1435 1 NOS1_RABIT	019132 oryctolagus
31	83	8.0	466 1 EM55_MOUSE	P70280 mus musculu
32	82.5	8.0	1010 1 SCA4_RICPA	094175 rickettsia
33	81.5	7.9	1174 1 ZO2_CANFA	095168 canis fami1

34	81	7.9	1319 1 MN1_HUMAN	Q10571 homo sapien
35	80	7.8	192 1 YL87_CAEEL	P46581 caenorhabd1
36	80	7.8	585 1 MPP3_HUMAN	Q13368 homo sapien
37	79.5	7.7	905 1 ZO3_MOUSE	094971 mus musculu
38	79.5	7.7	1012 1 SCA4_RICSL	094180 rickettsia
39	79.5	7.7	2254 1 CCAG_RAT	054898 rattus norv
40	78.5	7.6	341 1 CATL_DROME	Q95029 drosophila
41	78.5	7.6	1011 1 SCA4_RICCN	Q94382 rickettsia
42	78.5	7.6	1013 1 SCA4_RICRH	Q94381 rickettsia
43	78.5	7.6	1022 1 SCA4_RICCN	052658 rickettsia
44	78	7.6	581 1 PRLR_CEREL	028235 cervus elap
45	78	7.6	616 1 PRLR_RABIT	P14787 oryctolagus

ALIGNMENTS

RESULT 1
APYL_HUMAN STANDARD; PRT; 1616 AA.
ID APYL_HUMAN
AC 01396;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE APICAL-LIKE PROTEIN (APYL PROTEIN).
GN APYL.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=95315933; PubMed=7795590;
RA Schaffino V.M., Bassi M.T., Rugerli E.I., Renner A., Galli L.,
RA Ballabio A.;
RT "Cloning of a human homologue of the Xenopus laevis APX gene from the
RT ocular albinism type 1 critical region."
RT Hum. Mol. Genet. 4:373-382(1995).
RL [2]
RN SEQUENCE OF 56-1616 FROM N.A.
RP
RA Shen Y., Gabbas R.A.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
CC - TISSUE SPECIFICITY: ABUNDANT IN RETINA AND MELANOMA; ALSO IN
CC BRAIN, PLACENTA, LUNG, KIDNEY, PANCREAS.
CC - SIMILARITY: SOME TO XENOPUS LAEVIS APICAL PROTEIN (APX).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X83543; CAA58534.1;
DR EMBL: AC002365; AAC32592.1;
DR HSSP: 012959; 1PDR.
DR MIM: 300103;
DR InterPro: IPR000067; IGFBP.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00219; IGFBP; 1.
DR Pfam: PF00595; PDZ; 1.
DR SMART: SMO0228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
FT SIMILAR 54 112
FT TO MOUSE SYNTROPHIN-1 AND RABBIT
FT DYSTROPHIN-ASSOCIATED PROTEIN.
FT DOMAIN 150 153
FT POLY-SER.
FT DOMAIN 314 320
FT POLY-PRO.
FT DOMAIN 343 346
FT POLY-ALA.
FT DOMAIN 1065 1068
FT POLY-PRO.
SQ SEQUENCE 1616 AA; 176409 MM; 752406B5BC0B60A2 CRC64;

```

Query Match          28.2%; Score 290; DB 1; Length 1616;
Best Local Similarity 49.6%; Pred. No. 9.5e-19;
Matches 70; Conservative 18; Mismatches 43; Indels 10; Gaps 5;

QY 13 ATLNSMTATGGRVYILEAFLEGAPWGFTLKGLEHGEPLIISKVEGKADTLSSKLOA 72
DB 14 AEATRAADGRLV--EVQLSGAPWGFTLKGREHGEPLVITKIEGSKAAAV-DKLLA 70

QY 73 GDEVHNEVTLSSRKEAVSLVKGYKTLRLVRSLS-----PVTVLELDFDPQHPQ-R 126
DB 71 GDEIVGINDIGLQFRQEAICLVKGSHTKLKLVKRRSELGWRPHSWHATKFSDSHPELA 130

QY 127 MPPRTTSFSVSTADGRHWS 147
DB 131 ASPFTSTSGCPSWS-GRHHAS 150

RESULT 2
RIL_CHICK
ID RIL_CHICK STANDARD; PRT; 330 AA.
AC QSPW72;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN).
GN RIL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn;
RX MEDLINE=20379359; PubMed=10918612;
RA Fu S.-L., Maha A., Voigt P.K.;
RT "Identification and characterization of genes upregulated in cells transformed by v-Jun";
RL Oncogene 19:3537-3545(2000).
CC -!- FUNCTION: NEGATIVE REGULATOR OF CELL GROWTH.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC IONS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC EMBL; AF167295; AAD46655.1;
CC InterPro; IPR001781; LIM.
CC InterPro; IPR001478; PDZ.
CC Pfam; PF00412; LIM; 1.
CC Pfam; PF00595; PDZ; 1.
CC ProDom; PD000094; LIM; 1.
CC SMART; SM00132; LIM; 1.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00478; LIM_DOMAIN_1; 1.
CC PROSITE; PS50023; LIM_DOMAIN_2; 1.
CC PROSITE; PS50106; PDZ; 1.
CC LIM domain; Metal-binding; Zinc.
FT DOMAIN 8 84 PDZ.
FT DOMAIN 255 305 LIM.
SQ SEQUENCE 330 AA; 35838 MW; 2C0E0B404301E404 CRC64;

Query Match          13.0%; Score 134; DB 1; Length 330;
Best Local Similarity 29.4%; Pred. No. 3.7e-05;
Matches 45; Conservative 17; Mismatches 43; Indels 48; Gaps 6;

```

```

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90
DB 7 LRGPSFWGFRVLGVGKDFSTPLTISRINFGSKA--ALANLCPGDIILAINGESTEAMTHLE 64
QY 91 AVSLVKGYKTLRLVV-----RSLSPPV-----TVSLEFDP----- 121
DB 65 AQNKTKACVEQLLSVSRAEERSWSPPILEDGKAQAYRINIEPEPDNGPVGKRPMPHA 124
QY 122 -----OHPQMPBRTTSTFSVSTA 140
DB 125 AGGSPVDSRPALSLQHPQ--PSRPHASSADAA 155

RESULT 3
DLG2_RAT
ID DLG2_RAT STANDARD; PRT; 852 AA.
AC Q63622; Q62939; P70348;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSIN-110) (SYNAPTIC DENSITY PROTEIN PSD-93) (DISCS, LARGE HOMOLOG 2).
GN DLG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=963110881; PubMed=8755482;
RA Kim E., Cho K.O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins.";
RL Neuron 17:103-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96193770; PubMed=8625413;
RA Brenman J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E., Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F., Froehner S.C., Bret D.S.;
RA "Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
RL Cell 84:757-767(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Irie M., Hata Y., Takai Y.;
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: BELONGS TO A GUANYLATE KINASE-LIKE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC EMBL; U49049; AAB53243.1;
CC EMBL; U50717; AAC52643.1;
CC EMBL; U53368; AAB48562.1;
CC HSP; Q12959; LPDR.
CC InterPro; IPR000619; Guanylate_kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00625; Guanylate_kin; 1.
CC Pfam; PF00595; PDZ; 3.
CC Pfam; PF00018; SH3; 1.

```

```

RT  Chapsyn110 a member of the PSD-95 family of proteins." ;
RL  Neuron 17:103-113(1996). WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC  -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC  SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC  -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC  -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC  -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC  -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: U32376; AAB04949.1; .
DR  HSPF: Q12959; 1PDR.
DR  MIM: 603583;
DR  InterPro: IPR000619; Guanylate_kin.
DR  InterPro: IPR001478; PDZ.
DR  InterPro: IPR001452; SH3.
DR  Pfam: PF00625; Guanylate_kin; 1.
DR  Pfam: PF00595; PDZ; 3.
DR  Pfam: PF00018; SH3; 1.
DR  SMART: SM00072; Gukc; 1.
DR  SMART: SM00328; PDZ; 3.
DR  SMART: SM00326; SH3; 1.
DR  PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR  PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR  PROSITE: PS50106; PDZ; 3.
DR  PROSITE: PS50002; SH3; 1.
DR  SH3 domain; Repeat.
KW  DOMAIN
FT  DOMAIN 98 184 PDZ 1.
FT  DOMAIN 193 279 PDZ 2.

```

FT	DOMAIN	536	SH3.
FT	DOMAIN	680	GUANYLATE KINASE.
SO	SEQUENCE	870 AA;	89C83BA0619F6F59 CRC64;

	Query Match	12.5%;	Score 128.5;	DB 1;	Length 870;
	Best Local Similarity	28.2%;	Pred. No. 0.00041;		
	Matches	40;	Conservative	30;	Mismatches 47; Indels 25; Gaps
QY	10	KPSATINSTNTATKRG--RYLYLEAFLEGAGPFGTLGSLGEHPEPLISKEYEESGKADTLS	67		
		: : : : : : : : : : : : : : :			
Db	404	QPSMTLQRAVSLDEGEPRKVYLH---KGSTGLGFNFIVYG-EDGEGLFVSFLIAGGPAD-LTS	458		
		: : : : :			
QY	68	SKLQAGDEVVHIINEVTL--SSSRKEAVSLVKGSYKTRLVRVSLSPPVTVSLSEDFDPHQBR	122		
		: : : : : : : : : : : : : : :			
Db	459	GILRSGEDDITLVNNGIDLRASHEGOAAALAKGGCG--VYIIKQYPPEDVAR	507		
		: : : : :			

DB	Accession	Length	Score	E-value	Species
DB	508	FEAKIHDLREOMNMHSMSSGSG	529		

RESULT	5
PDLL_RAT	
ID	PDLL_RAT
AC	P52944;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL LIM DOMAIN PROTEIN 1) (ELF1N).
OS	PDLL1 OR CLP1 OR CLP36.
GN	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OX	NCBI_TaxID=10116;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

[1]
SEQUENCE FROM N.A.
STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
MEDLINE=96096533; PubMed=8522188;
Wang H., Harrison-Shostak D.C., Lemasters J.J.; Herman B.;
"Cloning of a rat cDNA encoding a novel LIM domain protein with high
homology to rat RIL";
Gene 165:267-271(1995).
CC -!- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
BRINGS OTHER PROTEINS TO THE CYTOSKELETON (BY SIMILARITY).
CC -!- SUBUNIT: INTERACTS WITH ALPHA-ACTININS 1, 2 AND 4 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN HEART, LUNG AND
LIVER, MODERATELY IN SPLEEN AND SKELETAL MUSCLE, AND AT EXTREMELY
LOW LEVELS (IF AT ALL) IN TESTIS AND BRAIN TISSUES.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U23769; AAA92046.1; .
InterPro; IPR001781; LIM.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF004112; LIM; 1.
DR Pfam; PF00595; PDZ; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; FALSE_NEG.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
DR PROSITE; PS50106; PDZ; 1.
CYtoskeleton; LIM domain; Metal-binding; Zinc.
DOMAIN 3 84 PDZ.
FT DOMAIN 258 308 LIM.
FT SEQUENCE 327 AA; 35525 MW; 4C3B5F579FBDA4D1 CRC64;

Query Match 12.3%; Score 127; DB 1; Length 327;
Best Local Similarity 32.6%; Pred. No. 0.00016;
Matches 43; Conservative 16; Mismatches 47; Indels 26; Gaps 5;

QY 32 LEGGAPWETLKGLHGEPELIISKVEGGKADTLSSKLAQGDEVVHN-EVTLSRRRE 90
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 8 LOGPGPWGFRVLGVGRDFEQLAISRVTGPSKA--AIANLCIGDLTAIDGETSSMTHLE 65

QY 91 AVSLVKGSYKTLRLVV-----RSLSPPVTVSLEFDPOHPORM-----PPR 130
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 66 AQNKIKGCVDNMTLVTSRSEQKINSPLVT---EEGKRHPYKMNLASPEQVELHGSAHNR 122

QY 131 TTTSFSVSTADG 142
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 123 SAMPTASPAPG 134

RESULT 6
PTN3_HUMAN
ID PTN3_HUMAN STANDARD; PRT; 913 AA.
AC P26045;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 3 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE H1) (PTP-H1).
GN PTPNS OR FTFPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

FT DOMAIN 670 913 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT SITE 842 842 BY SIMILARITY.
 SO SEQUENCE 913 AA; 104029 MM; 29A539ACDE2F1515 CRC64;

Query Match 11.9%; Score 122.5; DB 1; Length 913;
 Best Local Similarity 27.7%; Pred. No. 0.0016;
 Matches 33; Conservative 23; Mismatches 44; Indels 19; Gaps 4;

6 EDHFKDSATINSTATKGRY-----TYLEAFLEGAPMGFTLKGLHGEPLI 53
 Db 479 DDHRR--VTKGSTEDASQYCKNDNGDSYLVLRTPEDKGFENLKGVDKKRPLV 536
 QY 54 ISKVEEGKADTLSSKIQAGDEVHINEVTLSS-SRKAVALVGSYKT---LRLVVR 107
 Db 537 VSRINESPADTCIPKLNESGDIYLINGRDISEHTDQVYMFIKASRSHRELALVIR 595

RESULT 7
 ID RIL_RAT STANDARD; PRT; 330 AA.
 AC P36202;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN) (RIT-18) (H-REV18).
 GN RIL.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Liver;
 RX MEDLINE-95124724; Pubmed-7824279;
 RA Kless M., Scham B., Aguzzi A., Hajnal A., Klemenz R.,
 RA Schwarte-Waldhoff I., Schaefer R.,
 RT Expression of ril, a novel LIM domain gene, is down-regulated in
 RT Haas-transformed cells and restored in phenotypic revertants.
 RL Oncogene 10:61-68(1995).
 CC -1- TISSUE SPECIFICITY: DETECTED IN SEVERAL TISSUES, MOST PROMINENT IN
 CC BRAIN AND HEART OF ADULTS.
 CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
 CC IONS.

CC -1- SIMILARITY: CONTAINS 1 PDZ/DBP DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X76454; CA53992.1; -
 DR PIR: S39246; S39246.
 DR InterPro: IPR001781; LIM.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00412; LIM; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR ProDom: PD000094; LIM; 1.
 DR SMART: SM00132; LIM; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS00478; LIM_DOMAIN_1; 1.
 DR PROSITE: PS00023; LIM_DOMAIN_2; 1.
 DR PROSITE: PS0106; PDZ; 1.
 DR LIM domain; Metal-binding; zinc.
 KM LIM domain; Metal-binding; zinc.
 FT DOMAIN 1 84 PDZ.
 FT DOMAIN 255 305 LIM.
 SO SEQUENCE 330 AA; 35521 MM; 84BEB1E79ADA584 CRC64;

Query Match 11.8%; Score 121.5; DB 1; Length 330;

Best Local Similarity 32.0%; Pred. No. 0.00054;
 Matches 41; Conservative 15; Mismatches 51; Indels 21; Gaps 5;

QY 32 LEGAPMGFTLKGLHGEPLIISKVEEGKADTLSSKIQAGDEVHIN-EVTLSSRRE 90
 Db 7 LRGPSPMGFRVLGGRDFSAPLTISRVAHAGSKA--ALALCPGDSIOAINGESTELTLE 64
 QY 91 AVSLVGSYKTLRLVVRSLSPPTVS-----LEPDQPRQMPPTRTS--F 135
 Db 65 AQRIRKGDHDLTL---SVSRPNKNWPPSPNDKQAHRIHIDPEADGSPATSRSSIS 121
 QY 136 SVSTADOR 143
 Db 122 GISELNR 129

RESULT 8
 ID DLGI_RAT STANDARD; PRT; 911 AA.
 AC Q62656;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
 DE LARGE HOMOLOG 1).
 GN DLGI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE-95198112; Pubmed-7891172;
 RA Mueller B.M., Kistner U., Voh R.W., Cases-Langhoff C., Becker B.,
 RA Gundelfinger E.D., Garner C.C.;
 RT "Molecular characterization and spatial distribution of SAP97, a
 RT novel presynaptic protein homologous to SAP90 and the Drosophila
 RT discs-large tumor suppressor protein."
 RL J. Neurosci. 15:2354-2366(1995).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN
 CC CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO
 CC THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANES BETWEEN ADJACENT
 CC CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY
 CC SYNAPSES. IN OTHER BRAIN REGIONS, SAP97 IS FOUND IN AND ALONG
 CC BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS,
 CC BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A
 CC VARIETY OF EPITHELIAL CELLS.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DBP DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U14950; AAA79976.1; -
 DR HSSP: Q12959; IPDR.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.

DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS0106; PDZ; 3.
 DR PROSITE; PS0002; SH3; 1.
 KW SH3 domain; Repeat.
 FT DOMAIN 224 310 PDZ 1.
 FT DOMAIN 318 404 PDZ 2.
 FT DOMAIN 465 545 PDZ 3.
 FT DOMAIN 580 650 SH3.
 FT DOMAIN 721 911 GUANYLATE_KINASE.
 FT DOMAIN 527 530 POLY-ALA.
 SQ SEQUENCE 911 AA; 100570 MW; 18CEBD3DD0CAF8B CRC64;

Query Match 11.8%; Score 121; DB 1; Length 911;
 Best Local Similarity 25.8%; Pred. No. 0.0021;
 Matches 42; Conservative 30; Mismatches 55; Indels 36; Gaps 7;

QY 3 RTEDFHKPSATLNSNTATKGRYIYL-EAFL-----EGGAPWGFTLKGL 46
 DB 424 QSVNDHVPSSVLTGTPASPARISPKAVLGDDEITREPKVVLHKGSTGLGFINVGG- 482
 QY 47 EHEGLIISKVEEGKADTLSSKQAGDEVVHINEVTLS-SRKEAVSLVKGSKYKTLRLV 105
 DB 483 EDEGIGIFISFLLAGPAD-LSGELRKGDRISVNSVDLRAASHQAAAAAKNA----- 534
 QY 106 VRSLSPPVTVSLEFPDQHPQMPRP-----TRTSFVSVDADG 142
 DB 535 ----GQAVTIAQYRPEEYSRFEAKIHDLRETMNSSVSSGG 573

RESULT 9
 DLG3_MOUSE STANDARD; PRT; 849 AA.
 AC P70175;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102) (DISCS, LARGE HOMOLOG 3).
 GN DLG3 OR DLG3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Kohmura N., Makino S., Yagi T.;
 RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RL FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
 CC SUBUNIT NR2B
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D87117; BAA13249.1; -
 CC HSP; Q12959; LPDR.
 CC MGI; MGI:1888986; Dlg3.
 CC InterPro; IPR000619; Guanylate_kin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001452; SH3.

DR Pfam; PF00625; Guanylate_kin; 1.
 DR Pfam; PF00595; PDZ; 3.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00072; GUKC; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS0106; PDZ; 3.
 DR PROSITE; PS0002; SH3; 1.
 KW SH3 domain; Repeat.
 FT DOMAIN 149 235 PDZ 1.
 FT DOMAIN 244 330 PDZ 2.
 FT DOMAIN 404 484 PDZ 3.
 FT DOMAIN 519 589 SH3.
 FT DOMAIN 659 849 GUANYLATE_KINASE.
 SQ SEQUENCE 849 AA; 93482 MW; EF3EF2D7513538EE CRC64;

Query Match 11.6%; Score 119.5; DB 1; Length 849;
 Best Local Similarity 31.9%; Pred. No. 0.0027;
 Matches 37; Conservative 24; Mismatches 44; Indels 11; Gaps 6;

QY 10 KPSATLNSNTATKGRYIYLEAFLE-GGAPWGFTLKGLH-----GEP-LIISKVEEGSKA 63
 DB 133 EPSLSVN---GSDGMFKYEEIVLGRNGLGFSIAGGIDNPHVDPDPGIFITKIPGAA 189
 QY 64 DTLSSKQAGDEVVHINEVTLS-SRKEAVSLVKGSKYKTLRLVRSLSPPVTVSLE 118
 DB 190 -AMDGLGVNDCLRVNEVDVSEVVSRAVEALKEAGPVRLVRRRPPPTTIME 244

RESULT 10
 DLG3_RAT STANDARD; PRT; 849 AA.
 AC Q62936; P70547;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102) (PSD-95/SAP90 RELATED PROTEIN 1) (DISCS, LARGE HOMOLOG 3).
 GN DLG3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RC TISSUE=Brain;
 RX MEDLINE=96374358; PubMed=8780649;
 RA Mueller B.M., Kistner U., Kindler S., Chung W.J., Kuhlendahl S.,
 RA Fenster S.D., Lau L.-F., Veh R.W., Haganir R.L., Gundelfinger E.D.,
 RA Garner C.C.;
 RA "SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes in vivo."
 RT Neuron 17:255-265(1996).
 RL [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RA Irie M., Hata Y., Takai Y.;
 RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RL FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
 CC SUBUNIT NR2B
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D87117; BAA13249.1; -
 CC HSP; Q12959; LPDR.
 CC MGI; MGI:1888986; Dlg3.
 CC InterPro; IPR000619; Guanylate_kin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001452; SH3.

```

CC or send an email to license@lsb.slb.ch).
CC -----
CC EMBL: U50147; AAA93031.1;
CC DR EMBL: U5367; ABA48561.1;
CC DR HSP: 012959; 1PDR.
CC DR InterPro: IPR000619; Guanylate_kin.
CC DR InterPro: IPR001478; PDZ.
CC DR InterPro: IPR001452; SH3.
CC DR Pfam: PF00625; Guanylate_kin. 1.
CC DR Pfam: PF00595; PDZ; 3.
CC DR Pfam: PF00018; SH3; 1.
CC DR SMART: SM00072; GUKC; 1.
CC DR SMART: SM00228; PDZ; 3.
CC DR SMART: SM00326; SH3; 1.
CC DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
CC DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
CC DR PROSITE: PS50106; PDZ; 3.
CC DR PROSITE: PS50002; SH3; 1.
CC DR SH3 domain; Repeat; Alternative splicing.
CC SH3 domain; 149; 235; PDZ 1.
CC FT DOMAIN 149; 235; PDZ 1.
CC FT DOMAIN 244; 330; PDZ 2.
CC FT DOMAIN 404; 484; PDZ 3.
CC FT DOMAIN 519; 589; SH3.
CC FT DOMAIN 659; 849; GUANYLATE_KINASE.
CC FT VARSPLIC 627; 640; MISSING (IN SHORT ISOFORM).
CC SEQUENCE 849 AA; 93539 MW; 34DA9C46C7BB96DB CRC64;

Query Match 11.6%; Score 119.5; DB 1; Length 849;
Best Local Similarity 31.9%; Pred. No. 0.0027;
Matches 37; Conservative 24; Mismatches 44; Indels 11; Gaps 6;

OY 10 KPSATINSNTATKRYLYLEAFLE-GGAPMGFTLKGLEH---GEP-LIISKVEEGSKA 63
Db 133 EPSLSVN---GSDGMKEFEIYLERNGSLGFSINGIDNPDPGIFIKIIPGGA 189
OY 64 DFLSKLOAGDEVHINIVTSS-SKRAVSIVKSKYTLRLVYVSLPPTVSL 118
Db 190 -AMDRLGVNDCVLNVEVDYSEVHVSRAVLEKAGPVRLVVRROPPEITME 244

RESULT 11
ID PDL1_HUMAN STANDARD; PRT; 329 AA.
AC 000151;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
DE LIM DOMAIN PROTEIN 1) (ELFTN).
GN PDLIM1 OR CLIM1 OR CLP36.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=99145108; PubMed=10022510;
RA Lee C.Y., Wray M.M.Y., Garcia-Barcelo M., Tsui S.K.W., Fung K.P.,
RA "Characterization of the human 36-kDa carboxyl terminal LIM domain
RA protein (hCLIM1).";
RL J. Cell. Biochem. 72:279-285(1999).
RN [2]
RP INTERACTION WITH ALPHA-ACTININ 2.
RX MEDLINE=20320865; PubMed=10861853;
RA Kotaka M., Koshin S., Ngai S., Chan K., Lau Y., Lee S.M., Li H.Y.,
RA Ng E.K., Schaper J., Tsui S.K.W., Fung K.P., Lee C.Y., Wray M.M.Y.;
RA "Interaction of hCLIM1, an enigma family protein, with alpha-actinin
RA 2.";
RL J. Cell. Biochem. 78:558-565(2000).
RN [3]

```

```

RP INTERACTION WITH ALPHA-ACTININS 1 AND 4.
RX MEDLINE=20219155; PubMed=10753915;
RA Valentinus T., Luukko K., Mäkelä T.P.;
RT CLP-36 PDZ-LIM protein associates with nonmuscle alpha-actinin-1 and
RL alpha-actinin-4.
RL J. Biol. Chem. 275:11100-11105(2000).
CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON.
CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN THE HEART AND SKELETAL
CC MUSCLE, MODERATELY EXPRESSED IN THE SPLEEN, SMALL INTESTINE,
CC COLON, PLACENTA, AND LUNG. A LOWER LEVEL EXPRESSION IS SEEN IN
CC LIVER, THYMUS, KIDNEY, PROSTATE AND PANCREAS AND IS NOT FOUND IN
CC THE BRAIN, TESTIS, OVARY, AND PERIPHERAL BLOOD LEUKOCYTES.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@lsb-slb.ch).
CC -----
CC EMBL: U90878; AAC05580.1;
CC DR MIM: 605900;
CC DR InterPro: IPR001781; LIM.
CC DR InterPro: IPR001478; PDZ.
CC DR Pfam: PF00412; LIM; 1.
CC DR Pfam: PF00595; PDZ; 1.
CC DR Pfam: PF00094; LIM; 1.
CC DR ProDom: PD000094; LIM; 1.
CC DR SMART: SM00132; LIM; 1.
CC DR SMART: SM00228; PDZ; 1.
CC DR PROSITE: PS00478; LIM_DOMAIN_1; FALSE_NEG.
CC DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
CC DR PROSITE: PS50106; PDZ; 1.
CC DR Cytoskeleton; LIM domain; Metal-binding; Zinc.
CC FT DOMAIN 3 84
CC FT DOMAIN 260 310; LIM.
CC SEQUENCE 329 AA; 36171 MW; 6520F1A932CFA312 CRC64;

Query Match 11.4%; Score 117.5; DB 1; Length 329;
Best Local Similarity 36.3%; Pred. No. 0.0013;
Matches 37; Conservative 14; Mismatches 40; Indels 11; Gaps 4;

OY 32 LEGAPMGFTLKGLEHEPLIISKEEGKADTLSSKLOAGDEVHIN-EVTLSSSRKE 90
Db 8 LOGPGPWGFRVLGRDFEPLAISRYTPGSKA--ALANLCIGDVTALDGEINMTHTLE 65
OY 91 AVSLVGSYKTLRLVY-----RSLSPPTVSLIEDPQHPQRM 127
Db 66 AONRIKGTNDTLTVARSEHKVMSPLVT---EEGRHPYRM 104

RESULT 12
ID DGL1_HUMAN STANDARD; PRT; 904 AA.
AC 012959; 012958;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
DE LARGE HOMOLOG 1).
GN DGL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RESULT 13
DLG3_HUMAN
ID DLG3_HUMAN STANDARD; PRT; 817 AA.
AC Q92796; Q9ULI8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAPI02 (SYNAPSE-ASSOCIATED PROTEIN 102)
DE (NEUROENDOCRINE-DLG) (NE-DLG) (DISCS, LARGE HOMOLOG 3).
DE DLG3 OR KIAA1232.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Fetal brain;
MEDLINE=97332623; PubMed=9188857;
Makino K., Kuwahara H., Masuko N., Nishiyama Y., Morisaki T.,
Sasaki J., Nakao M., Kuwano A., Nakata M., Ushio Y., Saya H.;
"Cloning and characterization of NE-dlg: a novel human homolog of the
Drosophila discs large (dlg) tumor suppressor protein interacts with
the APC protein.";
ONCogene 14:2425-2433(1997).
[2]
SEQUENCE OF 330-817 FROM N.A.
TISSUE=Brain;
MEDLINE=20039619; PubMed=10574462;
Nagase T., Ishikawa K.-I., Kikuno R., Hirotsawa M., Nomura N.,
Ohata O.;
"Prediction of the coding sequences of unidentified human genes. XV.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
DNA Res 6:337-345(1999).
-!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
SUBUNIT NR2B (BY SIMILARITY).
-!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
-!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
-!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
EMBL; U49089; AABG1453.1;
EMBL; AB033058; BAA86546.1;
HSSP; Q12959; 1PDR.
MIM; 300189;
InterPro; IPR000619; Guanylate_kin.
InterPro; IPR001478; PDZ.
InterPro; IPR001452; SH3.
Pfam; PF00625; Guanylate_kin; 1.
Pfam; PF00595; PDZ; 3.
Pfam; PF00018; SH3; 1.
SMART; SM00072; GUKC; 1.
SMART; SM00228; PDZ; 3.
SMART; SM00326; SH3; 1.
PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
PROSITE; PS50106; PDZ; 3.
PROSITE; PS50002; SH3; 1.
SH3 domain; Repeat. 217 PDZ 1.
DOMAIN 130 226 311 PDZ 2.
DOMAIN 226 311 465 PDZ 3.
DOMAIN 379 465 568 SH3.
DOMAIN 503 568 SH3.

```


FT DOMAIN 3 84 PDZ.
FT DOMAIN 257 307 LIM.
SQ SEQUENCE 326 AA; 35717 MW; 5228220E59C1A90E CRC64;

Query Match 10.6%; Score 109; DB 1; Length 326;
Best Local Similarity 36.3%; Pred. NO. 0.0076;
Matches 37; Conservative 14; Mismatches 39; Indels 12; Gaps 5;

QY 32 LEGGAPWFTLKGLEHGEPLIISKVEGCKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90
|:|:||||| | | : :||| ||| | | | : | | | : | | | : | | |

Db 8 LOGPGWGFRLV-GKDFEQPLAISRTVPGSKA--AIANLCIGDLTAIDGEDTSSMTHLE 64
| : | | | : | | : | | | | | : | | | |

QY 91 AVSLVRGSKYKTLRLVY-----RSLSPPTVTVSLEFPDQHPQRM 127
| : | | | : | | : | | | | | : | | | |

Db 65 AQNKIKGCADNMTLTVSRSEQKWSPLVT---EGRKRHPYKM 103
| : | | | : | | : | | | | | : | | | |

Search completed: January 31, 2002, 21:06:28
Job time: 51 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:03:57 ; Search time 72.88 seconds
(without alignments)
381.336 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MKRTEDFHKPSATLNSNTA.....CIYPQHNTGKNEKRAVLSR 190

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_17:*

1: sp-archaea:*

2: sp-bacteria:*

3: sp-fungi:*

4: sp-human:*

5: sp-invertebrate:*

6: sp-mammal:*

7: sp-mhc:*

8: sp-organellar:*

9: sp-phage:*

10: sp-plant:*

11: sp-rodent:*

12: sp-virus:*

13: sp-vertebrate:*

14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	426	41.4	1986	11	Q9GXNO mus musculu
2	271.5	26.4	1502	4	Q9D1L8 Q9H18 homo sapien
3	143.5	13.9	1454	4	Q9H4G2 Q9H4G2 homo sapien
4	138.5	13.5	298	13	Q9PU46 Q9PU46 gallus gall
5	138.5	13.5	315	13	Q9PU47 Q9PU47 gallus gall
6	136	13.2	470	4	Q9Y4Z3 Q9Y4Z3 homo sapien
7	135.5	13.2	661	11	Q9JKS3 Q9JKS3 mus musculu
8	135.5	13.2	723	11	Q9JKS4 Q9JKS4 mus musculu
9	135.5	13.2	723	11	Q9VWH2 Q9VWH2 mus musculu
10	133.5	13.0	171	5	Q17458 Q17458 schistosoma
11	132.5	12.9	234	11	Q9D2Q1 Q9D2Q1 mus musculu
12	131	12.7	283	4	Q9Y4Z5 Q9Y4Z5 homo sapien
13	131	12.7	288	11	Q9R0Z1 Q9R0Z1 mus musculu
14	131	12.7	288	11	Q9VWH1 Q9VWH1 mus musculu
15	131	12.7	596	4	Q60705 Q60705 homo sapien
16	131	12.7	617	4	Q9Y4Z4 Q9Y4Z4 homo sapien
17	131	12.7	1278	4	Q60434 Q60434 homo sapien
18	131	12.7	1455	4	Q60510 Q60510 homo sapien
19	130.5	12.7	138	11	Q9D130 Q9D130 mus musculu

20	130.5	12.7	493	5	Q9W443 Q9W443 drosophila
21	130.5	12.7	734	4	Q75112 Q75112 homo sapien
22	129.5	12.6	207	11	Q9Z252 Q9Z252 rattus norv
23	128.5	12.5	207	4	Q9HAP6 Q9HAP6 homo sapien
24	128.5	12.5	400	5	Q9V7F2 Q9V7F2 drosophila
25	128	12.4	39	4	Q9H4T0 Q9H4T0 homo sapien
26	128	12.4	435	4	Q14250 Q14250 homo sapien
27	127.5	12.4	316	11	Q70209 Q70209 mus musculu
28	127	12.3	337	11	Q9QYN0 Q9QYN0 mus musculu
29	127	12.3	591	11	Q9QYN2 Q9QYN2 mus musculu
30	126.5	12.3	207	4	Q88951 Q88951 mus musculu
31	126.5	12.3	403	4	Q9UPC3 Q9UPC3 homo sapien
32	126.5	12.3	552	4	Q9UM04 Q9UM04 homo sapien
33	126.5	12.3	652	4	Q9Y6N9 Q9Y6N9 homo sapien
34	126	12.2	195	5	Q9VBY7 Q9VBY7 drosophila
35	125.5	12.2	327	11	Q99K93 Q99K93 mus musculu
36	125.5	12.2	329	4	Q9BP29 Q9BP29 homo sapien
37	124.5	12.1	153	4	Q9BXB8 Q9BXB8 homo sapien
38	124.5	12.1	214	11	Q9CRA2 Q9CRA2 mus musculu
39	124.5	12.1	457	4	Q9NR12 Q9NR12 homo sapien
40	124.5	12.1	521	4	Q9UM17 Q9UM17 homo sapien
41	124	12.1	591	11	Q62920 Q62920 rattus norv
42	123.5	12.0	548	11	Q9ES65 Q9ES65 mus musculu
43	123.5	12.0	910	11	Q9ES64 Q9ES64 mus musculu
44	122.5	11.9	189	4	Q9BYP4 Q9BYP4 homo sapien
45	122	11.9	362	11	Q70208 Q70208 rattus norv

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	1986 AA.
Q9GXNO	Q9GXNO			
AC	Q9GXNO			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	PDZ DOMAIN ACTIN BINDING PROTEIN SHROOM.			
GN	SHRM.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=2005594; PubMed=10589677;			
RX	Hildebrand J.D., Soriano P.;			
RA	"Shroom, a PDZ domain-containing actin-binding protein, is required			
RT	for neural tube morphogenesis in mice."			
RL	Cell 99:485-497(1999).			
DR	EMBL: AF199421; AAF13269.1; -			
DR	HSSP: P29476; ICAY.			
DR	MGD: MGI:1351655; Shrm.			
DR	InterPro: IPR001478; PDZ.			
DR	Pfam: PF00595; PDZ; 1.			
DR	SMART: SM00228; PDZ; 1.			
DR	PROSITE: PS50106; PDZ; 1.			
DR	SEQUENCE 1986 AA; 215261 MW; 18C5EFCB772EABF CRC64;			

Query Match	41.4%; Score 426; DB 11; Length 1986;
Best local similarity	46.3%; Pred. No. 6.3e-30;
Matches 101; Conservative 24; Mismatches 57; Indels 36; Gaps 6;	
QY	2 MKRTEDFHKPSATLNSNTKRYIYIEFLGGAHPWFTLKGLGHEGPLIISVEGG 61
DB	1 MKTPETLEPSPATPNPSRPTERFYILELLEGGAPWFTLKGLGHEGPLIISKEEG 60
QY	62 KADITSSKIQAGDEVHINEVTLLSSRKRAVSLVKGSKYTLRLVYR-----SLSPV 113
DB	61 KADYSSGQIAGDEVHINEVTLLSSPRRAVSLVKGSKYTLRLVYRVDYCAAPGADPGT 120

```
QY 114 TVSLFED-----POHPQ-----RMPRTTSFSVSTADGRHEWSCRPVWVWWSRPRT 161
DB 121 SKSLSELLTCSQPRKATSGVYKLRKQKCS---EPATRPWSWHT-----TKGETQPD 173
QY 162 WAARWPQKGCYPTQHT-----CRNFKRAYLSR 190
DB 174 VSMQISQGTMPGPHWQSHSSSTSDLSNYDHAYLRR 211

RESULT 2
ID Q9ULL8 PRELIMINARY; PRT; 1502 AA.
AC Q9ULL8;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE KIAA1202 PROTEIN (FRAGMENT).
GN KIAA1202.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033028; BAA86516.1; -.
DR HSSP; P31016; 1BE9.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 1502 AA; 165950 MW; DC050C9C69B78AD9 CRC64;

Query Match 26.4%; Score 271.5; DB 4; Length 1502;
Best Local Similarity 56.5%; Pred. No. 5.4e-16;
Matches 52; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

QY 23 GRYILEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEV 82
DB 10 GSFQYVPVQLGGAPWGFTLKGLEHGEPLIYSKIEDGGKA-ALSKMRTGDELVNINGT 68
QY 83 TLSSSRKEAVSLVKGSKYKTLRLVVRSLSPVVT 114
DB 69 PLYGSRQREALILIKGSRILKILVRRNAPVS 100

RESULT 3
QYH4G2
ID Q9H4G2 PRELIMINARY; PRT; 1454 AA.
AC Q9H4G2;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE BAI19E20.1 (KIAA1202 PROTEIN) (FRAGMENT).
GN BAI19E20.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson S.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121865; CAC08472.1; -.
DR InterPro; IPR001478; PDZ.
```

```
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 1454 AA; 160615 MW; 7931571D7DA80DE8 CRC64;

Query Match 13.9%; Score 143.5; DB 4; Length 1454;
Best Local Similarity 50.0%; Pred. No. 0.00024;
Matches 29; Conservative 15; Mismatches 13; Indels 1; Gaps 1;

QY 57 VEEGKADTLSSKLOAGDEVVHINEVTLSSSRKEAVSLVKGSKYKTLRLVVRSLSPVVT 114
DB 1 IEDGGKA-ALSKMRTGDELVNINGTPLYGSRQREALILIKGSRILKILVRRNAPVS 57

RESULT 4
QYPU46
ID Q9PU46 PRELIMINARY; PRT; 298 AA.
AC Q9PU46;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SKELETAL MUSCLE ISOFORM (SKALP)
DE (FRAGMENT).
GN ALP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436131; PubMed=10506181;
RA Ponies P., Macalima T., Beckerle M.C.;
RT "Purification and characterization of an alpha-actinin-binding PDZ-LIM
RT protein that is up-regulated during muscle differentiation.";
RL J. Biol. Chem. 274:29242-29250(1999).
DR EMBL; AJ249219; CAB53971.1; -.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
FT NON_TER 298
SQ SEQUENCE 298 AA; 32094 MW; 27D6F24FC6141E61 CRC64;

Query Match 13.5%; Score 138.5; DB 13; Length 298;
Best Local Similarity 31.7%; Pred. No. 9.2e-05;
Matches 38; Conservative 21; Mismatches 38; Indels 23; Gaps 5;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEV-TLSSSRKE 90
DB 7 LFGAPWGFRSLGGIDFNQPLIITRITPGSKAST--ANLCPGDIIVAINGLSTENHTND 64
QY 91 AVSLVKGSKYKTLRLVVRSL-----SP-----PVTVSLEFPDQ-----HPQRMPPR 130
DB 65 AGERIAAAHQSLRIKRAETKLNSQVSDGKANYKINLEAFQDINFEHHRIRPK 124

RESULT 5
QYPU47
ID Q9PU47 PRELIMINARY; PRT; 315 AA.
AC Q9PU47;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SMOOTH MUSCLE ISOFORM (SMALP).
GN ALP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
```



```
QY      32 LEGAPMFGFTLKGLEHGEPLITISKVEEGKADLLSKLAGEDEVYINNEY-TLSSSRKE 90
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      7 LPGPAPMGFRLSGIDFNOPLIITRIRTPGSKAST--ANLCRGDIIVAINGSTENMTND 64
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      91 AVSLVKGSYKTLLRV-----RSLSPP-----PTVSLLEFPDQ-----HPQRMP 128
        | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      65 AOERKKAHAHOLSLIERAETKLMSPOVSDEGKANPYTINIIEARPEQEKPIGTAAHNRAQ 128
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      129 P 129
        |
Db      125 P 125
```

```

RESULT      6
O9Y4Z3      PRELIMINARY;      PRT;      470 AA.
ID      O9Y4Z3
AC      O9Y4Z3:
DT      01-NOV-1999 (TREMBLrel. 12, Created)
DT      01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      ZASP PROTEIN (FRAGMENT).
GN      ZASP.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE-SKELETAL MUSCLE, AND HEART;
RA      Trevisan G., Pallavicini A., Formentin E., Comella A., Ievoliella C.,
RA      Trevisan S., Portolotto G., Scannapieco P., Salamon M., Mouly G.,
RA      Valle G., Lanfranchi G.,
RT      "ZASP: a new 2-band alternatively spliced PDZ-motif protein.";
RL      J. Cell Biol. 0:0-0(0).
CC      -1-SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
CC      IONS.
DR      EMBL: AJ133768; CAB46729.1; -.
DR      HSPR: Q05158; IOLI.
DR      InterPro: IPR003006; IG_MHC.
DR      InterPro: IPR01781; LIM.
DR      InterPro: IPR001478; PDZ.
DR      InterPro: IPR002965; P_Rich_extensn.
DR      Pfam: PF00412; LIM; 3.
DR      Pfam: PF00595; PDZ; 1.
DR      PRINTS: PR01217; PRICHXENSN.
DR      ProDom: PD000094; LIM; 3.

```

Query Match	13.2%;	Score 136;	DB 4;	Length 470;
Best Local Similarity	29.8%;	Pred. No. 0.00026;		
Matches 53;	Conservative 20;	Mismatches 61;	Indels 44;	Gaps 8

```

QY      32 LEGAPMGFTLKGLGHEHPEPLISKVEEGKADPLSSKLAGDEPVHINEY-ETSSSRKE 90
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7  LTGCPMGFTFRLOGKDEPMPLTISRTPGSA--AQSLGSGDLVWALDGVNDITMHL 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      91 AVLTVKSGYKTLRLY-----VRSISPVTLSLEF-----DPQHPQMPRTR 132
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      65 AQNKTKSASYNLSLTLQSKRPPISTIAPVVOQPLPIPIHQKVYVSNPANAASABATH 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      133 TSFS-----VSTADGRHEWSCRPWVYKWSBPRPTWAARWPOKICIT-PTQH 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      125 TSYSGPAAPAPKPRVVTIASIR-----PSVQOPVASTYS--PSPGANYSPTEPY 172
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT	7	
	09JKS3	
ID	09JKS3	PRELIMINARY: PRT: 661 AA.
AC	09JKS3	
DT	01-OCT-2000 (TEMBLrel. 15, Created)	
DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)	
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)	
DE	ORACLE 2 PROTEIN (FRAGMENT).	
OS	Mus musculus (mouse)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
OX	NCBI_TaxId=10090;	

RP SEQUENCE FROM N.A.
RX MEDLINE-20193503: PubMed-10727866;
RA Passier R., Richardson J.A., Olson E.N.;
RT "Oracle, a novel PDZ-LIM domain protein expressed in heart and
skeletal muscle.";
RL Mech. Dev. 92:277-284(2000).
CC -1- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
IONS.
CC EMBL: AF228058; AAF33848.1; -.
DR InterPro: IPR000345; Cyt_c_heme_bind.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR001781; LIM.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR002965; P_rich_extensn.
DR Pfam: PF00412; LIM; 3.
DR Pfam: PF00595; PDZ; 1.
DR PRINTS: PRO1217; PRICHEXTENSN.
DR PRODOM: PD000094; LIM; 3.
DR SMART: SM00132; LIM; 3.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE: PS00478; LIM_DOMAIN_1; 2.
DR PROSITE: PS50023; LIM_DOMAIN_2; 3.
DR PROSITE: PS50106; PDZ; 1.
FW LIM motif; metal-binding; zinc.
KW NON_TER 661 661
SQ SEQUENCE 661 AA; 70750 MM; 5DFFEEF78D45532F CRC64;

Query Match	13.28;	Score 135.5;	DB 11;	Length 661
Best Local Similarity	31.78;	Pred. No. 0.00048;		

Db 125 EARSP 130

RESULT	9	
Q9WH2		
ID	Q9WH2	PRELIMINARY;
AC	Q9WH2;	PRT; 723 AA.
DT	01-NOV-1999	(TREMBLrel. 12, Created)
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)

OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX	NCBI_TaxId:10090;
RI	[1]
RN	SEQUENCE FROM N.A.
RP	MEDLINE-99321910; PubMed-10391924;
RX	Zhou Q., Ruiz-Lozano P., Martone M.E., Chen J.;
RA	"Cypher, a Striated Muscle-restricted PDZ and LIM Domain-containing
RT	protein, Binds to alpha-Actinin-2 and Protein Kinase C.";
RL	J. Biol. Chem. 274:13907-19813(1999).
RR	[2]
RN	SEQUENCE FROM N.A.
RP	Chen J., Zhou Q.;
RA	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

Query Match	13.2%;	Score 135.5;	DB 11;	Length 723;	
Best Local Similarity	31.7%;	Pred. No. 0.00053;			
Matches	40;	Conservative 19;	Mismatches 58;	Indels 9;	Gaps
QY	32	LEGGAPWGFTLKGLEHGPELLISKVEEGKADTLLSSKLOAGDEVVHINEV-TLSSSRYE	90		
Db	7	LTGPGWGFRLQGGKDFNNPLTISRITPGSKA--AQSQLSQGDLVVVAIDGVNTDTMTHLE	64		
QY	91	AVSLVKGSKYKTLRLVVRSLSPPVTVSLEFPDQH-----PQRMPPRTRTSFSVSTADGRH	144		
Db	65	AQNKIKASYNLSLTLQSKREPIPISTTAPPIQSLPLVPVPHQKDPALDINGSIATPSPSP	124		
QY	145	EWSCR	150		
Db	125	EARSP	130		
RESULT	10				
Q17458					
ID	Q17458	PRELIMINARY;	PRT;	171 AA.	
AC	Q17458;				
DT	01-JAN-1998	(TremBLrel. 05, Created)			
DD	01-JAN-1998	(TremBLrel. 05, Last sequence update)			
DT	01-JUN-2001	(TremBLrel. 17, Last annotation update)			

DE HYPOTHETICAL 19.0 KDA PROTEIN (FRAGMENT).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Rhabdophora; Neodermata;
 OC Trematoda; Digenea; Strigoida; Schistosomatidae; Schistosomatidae;
 OC Schistosoma.
 NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PUERTO RICAN;
 RA Harrop R., Wilson R.A.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF030965; AAB86566.1; -
 DR HSSP: P31016; 1BFE.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50106; PDZ; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 171 AA; 19000 MW; 69685F4108598063 CRC64;
 Query Match 13.0%; Score 133.5; DB 5; Length 171;
 Best Local Similarity 33.8%; Pred. No. 0.00013;
 Matches 47; Conservative 24; Mismatches 47; Indels 21; Gaps 8;
 QY 4 TTEDFH-----KPSATLNSN---TATKG---RYIYLEAFLEGGAPWGFTLKGLEHGE 50
 Db 34 TTVYDNGSEEVKASATATYAAFAASEGHAPRYELPKTNEG---LGFNVWGKKEQNS 90
 QY 51 PLISYVEEGKADTLSSKLOAGEVHINEVTLSSSRKE-AVSLVKGSKYTLRVSL 109
 Db 91 PIYISRMKPGGVADR-HGGLKRGDQLSVNGISVESEHHEKAVELLLKAGVQLVLR-Y 148
 QY 110 SPPTVSLF--FDPOHQR 126
 Db 149 TPRIIEEMAEARFQKOKARR 167
 RESULT 11
 Q9D201 PRELIMINARY; PRT; 234 AA.
 ID Q9D201
 AC 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VERTEBRATE HOMOLOG OF C. ELEGANS LIN-7 TYPE 2.
 GN LIN-7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPI;
 RL MEDLINE=21083660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
 Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Glassl C., King B., Kochiya H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombarts P.,
 McDermott P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasakhi H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
 Suzuki H., Taya-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,

RA Hayashizaki Y.;
 RL "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK019299; BAB31655.1; -
 DR MGD: MG1:1330858; Vell12.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50106; PDZ; 1.
 SQ SEQUENCE 234 AA; 25732 MW; 76919D8B1C29E53E CRC64;
 Query Match 12.9%; Score 132.5; DB 11; Length 234;
 Best Local Similarity 32.9%; Pred. No. 0.00024;
 Matches 47; Conservative 25; Mismatches 44; Indels 27; Gaps 8;
 QY 13 ATLNSNTATKG---RYIYLEAFLEGGAPWGFTLKGLEHGEPLISYVEEGKADTLSS 68
 Db 77 ATVAFAFASGHAHPRVVELPKTDEG---LGFNVWGKKEQNSPIYISHVIFGVDAR-HG 132
 QY 69 KLOAGEVHINEVTLSSSRKE-AVSLVKGSKYTLRVSLSPPTVSLF--FD----- 120
 Db 133 GLKRGDQLSVNGISVESEHHEKAVELLLKAGSVQLVLR-YTPRVLEEMAEAREKRSR 191
 QY 121 ---PQHPO-----RMPDPTRT 133
 Db 192 RRRPAPQLHVLGVSRKLSQTRT 214
 RESULT 12
 Q9Y4Z5 PRELIMINARY; PRT; 283 AA.
 ID Q9Y4Z5
 AC 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ZASP PROTEIN (FRAGMENT).
 GN ZASP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKELTAL MUSCLE, AND HEART;
 RA Faulkner G., Pallavicini A., Portentun E., Comelli A., Ievoliella C.,
 Trevisan S., Bortoletto G., Scannapieco P., Salamon M., Mouty G.,
 Valle G., Lanfranchi G.;
 RT "ZASP: a new Z-band alternatively spliced PDZ-motif protein.";
 RL J. Cell Biol. 0:0-0(0).
 DR EMBL: AJ133766; CAB46727.1; -
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR SMART: SM00228; PDZ; 1.
 FT NON_TER 1
 SQ SEQUENCE 283 AA; 30791 MW; F6384C88E71767E9 CRC64;
 Query Match 12.7%; Score 131; DB 4; Length 283;
 Best Local Similarity 30.1%; Pred. No. 0.00042;
 Matches 46; Conservative 25; Mismatches 74; Indels 8; Gaps 5;
 QY 32 LBSGAWGTTLKGLEHGEPLISYVEEGKADTLSSKLOAGEVHINEV-TLSSSRKE 90
 Db 7 LSGPGWGRLOGGKDFNPLTRITPSKA--AOSQSGGLVAIDGVNDTWTLE 64
 QY 91 AVSLVGSYKTLRVSLSPPTVSLFDP--OHPOMPDPTRTSPVSSTADGRHEMSCR 149
 Db 65 AOKKITSASVNLSTLOKSKRPISITAPPVOTPLPVRPHQKVYVN-SPANADYGERFN 123
 QY 150 PPVVK---WMSPPRTWAARMPKGCITYPROHNT 179

Chen J., Zhou Q.,
Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF114379; AAD42951.2; -
MGD; MGI:134412; zasp.
InterPro; IPR001478; PDZ.
InterPro; IPR000863; Sulfotransferase.
Pfam; PF00595; PDZ; 1.
ProDom; PD001218; Sulfotransfer; 1.
SMART; SM00228; PDZ; 1.
PROSITE; PS50106; PDZ; 1.
SEQUENCE 288 AA; 31427 MW; 65B11275409291B6 CRC64;

Query Match 12.7%; Score 131; DB 11; Length 288;
Best Local Similarity 30.1%; Pred. No. 0.00043; Indels 8; Gaps 5;
Matches 46; Conservative 25; Mismatches 74; Indels 8; Gaps 5;

QY 32 LEGGAPWGFTLKGLEHGEPLLIISKVEGGKADTLSSKLQAGDVEVHINEV-TLSSSRKE 90
DB 7 LTGPGWPGFLOGGKDFNMPLTISRITPGSKA--AQSQLSGDLVAIDGVTNTDTMTHLE 64

QY 91 AVSLVKGSKYTLRLVRSLSPPVTVSLEFPD-QHPQRPMPTRTSFSYSTADGRHNSCR 149
DB 65 AQNKIKSASYNLSLTLLQKSRPISTAPTQISPLEVIPHQKV-VANSPANADQIERFN 123

QY 150 PPMVK--NWSRPRTWAARWPKGKCIYPTQHT 179
DB 124 PSVLKUSALSTHRPIEVKGLGGKATIIHAQYNT 156

RESULT 15
O60705 PRELIMINARY; PRT; 596 AA.
AC O60705;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LIM PROTEIN.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RL Yano K., Oda T., Kondo M., Ueki N.;
LA Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC EMBL; AF061258; AAC15767.1; -
DR InterPro; IPR001781; LIM.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00412; LIM; 3.
DR Pfam; PF00595; PDZ; 1.
DR ProDom; PD000094; LIM; 3.
DR SMART; SM00132; LIM; 3.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 2.
DR PROSITE; PS0023; LIM_DOMAIN_2; 3.
DR PROSITE; PS50106; PDZ; 1.
KW LIM motif; Metal-binding; Zinc.
SQ SEQUENCE 596 AA; 64028 MW; 227652B643A1426D CRC64;

Query Match 12.7%; Score 131; DB 4; Length 596;
Best Local Similarity 29.9%; Pred. No. 0.0011;
Matches 41; Conservative 28; Mismatches 58; Indels 10; Gaps 5;

QY 32 LEGGAPWGFTLKGLEHGEPLLIISKVEGGKADTLSSKLQAGDVEVHINEV-TLSS-SRKE 90
DB 8 LVGPAPWGFTLOGGKDFNMPLTISLKDGGKA--AQANVRIGDGVLLSIDGINAOGMTHLE 65

QY 91 AVSLVKGSKYTLRLVRSLS-----PPVTVSLEFPDQHPQRPMPTRTTSFVSVDGRHE 145

Fri Feb 1 08:56:27 2002

us-09-641-831-2.rspt

Page 7

Db 66 AONKIKGCTGSLNMTLORASAPRPEPVQ-KGEPRKVVKPVITSPAVSKVTSTNMA 124

QY 146 WSCR-PWTKWSDRPT 161

Db 125 YNKAPRPRGSVSPKVT 141

Search completed: January 31, 2002, 21:05:34
Job time: 97 sec

XX Novel human polynucleotides isolated from human mammary gland cDNA
PT library, encodes novel human proteins which are useful in diagnosis,
PT drug screening, clinical trial monitoring or treating behavioural
PT disorders -
PS
XX Claim 3; Page 25; 29pp; English.
XX The sequence encodes a novel human protein, NHP#2, which shares
CC structural motifs with human APXL protein. Nucleotide constructs
CC encoding functional NHPs are used in gene therapy approaches for the
CC modulation of NHP expression. NHP oligonucleotides can be used as
CC hybridisation probes for screening libraries and assessing NHP gene
CC expression patterns. Also, labeled NHP nucleotide probes can be used to
CC screen a human genomic library. The NHP nucleotide sequences are also
CC useful in drug screening techniques for treating symptomatic or
CC phenotypic manifestations of perturbing the normal function of NHP in the
CC body. Examples of such manifestations may include obesity and high blood
CC pressure.
XX
XX Sequence 327 BP; 91 A; 76 C; 96 G; 64 T; 0 other;

Query Match 100.0%; Score 327; DB 22; Length 327;
Best Local Similarity 100.0%; Pred. No. 1.3e-95;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgatgagacactgaagactccacaaagccttagtgccacattaaacttaaacacgccc 60
Db 1 atgatgagacactgaagactccacaaagccttagtgccacattaaacttaaacacgccc 60
Qy 61 accaaggaaggtacattattctgagcattctcctgagggagagcctccctggggtttt 120
Db 61 accaaggaaggtacattattctgagcattctcctgagggagagcctccctggggtttt 120
Qy 121 actctaaagggtgctgagcagcagagacatttaattctcttaaggtcgaaagagg 180
Db 121 actctaaagggtgctgagcagcagagacatttaattctcttaaggtcgaaagagg 180
Qy 181 ggcaagcagacacccctgagctccaaactgcaggctggggatgaggtgtgcacatcaat 240
Db 181 ggcaagcagacacccctgagctccaaactgcaggctggggatgaggtgtgcacatcaat 240
Qy 241 gaggtgactctgagcagctccagaaaggagcaggttccctgggtgaaaggtatcctacaag 300
Db 241 gaggtgactctgagcagctccagaaaggagcaggttccctgggtgaaaggtatcctacaag 300
Qy 301 accctcaggctgtagtgcagttga 327
Db 301 accctcaggctgtagtgcagttga 327

RESULT 2
AAS00041
ID AAS00041 standard; cDNA; 402 BP.
XX
AC AAS00041;
XX
DT 11-MAY-2001 (first entry)
XX
DE Human cDNA encoding novel human protein, NHP#3.
XX
KW Human: novel human protein; NHP#3; gene therapy; drug screening;
KW obesity; high blood pressure; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..402
FT /tag= a
FT /product= "NHP#3"
XX
PN WO200114422-A1.

XX PD 01-MAR-2001.
XX XX
XX 18-AUG-2000; 2000WO-US22815.
XX PF
XX 24-AUG-1999; 99US-0150511.
XX PR
XX (LEXI-) LEXICON GENETICS INC.
XX PA
XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
XX PI
XX WPI; 2001-218430/22.
XX DR P-PSDB; AAU00033.
XX DR
XX Novel human polynucleotides isolated from human mammary gland cDNA
PT library, encodes novel human proteins which are useful in diagnosis,
PT drug screening, clinical trial monitoring or treating behavioural
PT disorders -
XX
XX Claim 5; Page 25; 29pp; English.
XX The sequence encodes a novel human protein, NHP#3, which shares
CC structural motifs with human APXL protein. Nucleotide constructs
CC encoding functional NHPs are used in gene therapy approaches for the
CC modulation of NHP expression. NHP oligonucleotides can be used as
CC hybridisation probes for screening libraries and assessing NHP gene
CC expression patterns. Also, labeled NHP nucleotide probes can be used to
CC screen a human genomic library. The NHP nucleotide sequences are also
CC useful in drug screening techniques for treating symptomatic or
CC phenotypic manifestations of perturbing the normal function of NHP in the
CC body. Examples of such manifestations may include obesity and high blood
CC pressure.
XX
XX Sequence 402 BP; 110 A; 91 C; 117 G; 84 T; 0 other;

Query Match 98.8%; Score 323; DB 22; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.7e-94;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgatgagacactgaagactccacaaagccttagtgccacattaaacttaaacacgccc 60
Db 1 atgatgagacactgaagactccacaaagccttagtgccacattaaacttaaacacgccc 60
Qy 61 accaaggaaggtacattattctgagcattctcctgagggagagcctccctggggtttt 120
Db 61 accaaggaaggtacattattctgagcattctcctgagggagagcctccctggggtttt 120
Qy 121 actctaaagggtgctgagcagcagagacatttaattctcttaaggtcgaaagagg 180
Db 121 actctaaagggtgctgagcagcagagacatttaattctcttaaggtcgaaagagg 180
Qy 181 ggcaagcagacacccctgagctccaaactgcaggctggggatgaggtgtgcacatcaat 240
Db 181 ggcaagcagacacccctgagctccaaactgcaggctggggatgaggtgtgcacatcaat 240
Qy 241 gaggtgactctgagcagctccagaaaggagcaggttccctgggtgaaaggtatcctacaag 300
Db 241 gaggtgactctgagcagctccagaaaggagcaggttccctgggtgaaaggtatcctacaag 300
Qy 301 accctcaggctgtagtgcag 323
Db 301 accctcaggctgtagtgcag 323

RESULT 3
AAS00039
ID AAS00039 standard; cDNA; 573 BP.
XX
AC AAS00039;
XX
DT 11-MAY-2001 (first entry)
XX

ID AAC63964 standard; DNA; 2566 BP.
XX
AC AAC63964;

DT 09-FEB-2001 (first entry)

XX Methylophilus methylophilus diaminopimelate decarboxylase gene.

XX Lysine biosynthesis; diaminopimelate decarboxylase;

KM methanol carbon source; enhanced enzyme activity;

XX L-lysine acid production; transgenic bacterium; Methylophilus; ds.

OS Methylophilus methylophilus.

PN WO200061723-A1.

PD 19-OCT-2000.

XX 07-APR-2000; 2000MO-JF02295.

PR 09-APR-1999; 99JP-0103143.

PR 16-JUN-1999; 99JP-0169447.

PR 24-DEC-1999; 99JP-0368097.

XX (AJIN) AJINOMOTO CO INC.

PI Gunji Y, Yasueda H, Sugimoto S, Tsujimoto N, Shimaoka M, Miyata Y;

PI Oda M;

DR WPI: 2000-672679/65.

DR P-PSDB; AAB29520.

XX Efficient production of L-lysine by fermenting

PT transformed Methylophilus bacterium with enhanced dihydrodipicolinate

PT synthase and aspartokinase activities or casamino acid requirement

PT using methanol

XX Claim 25; Page 77-80; 92pp; Japanese.

XX The invention relates to a transformed bacterium of the genus

CC Methylophilus which is capable of producing an L-amino acid efficiently

CC using methanol as the main carbon source. The biosynthetic pathway for

CC the desired amino acid has enhanced activity in the transformed

CC bacterium. In particular, the amino acid that is produced is lysine,

CC and the L-lysine biosynthetic enzymes are not subject to feedback

CC inhibition by lysine. The invention provides genes (AAC63958-C63964)

CC encoding the Methylophilus methylophilus lysine biosynthetic enzymes

CC aspartokinase (AAB29516), aspartic semialdehyde dehydrogenase

CC (AAB29517), dihydrodipicolinate synthase (AAB29518), dihydrodipicolinate

CC reductase (AAB29519), and diaminopimelate decarboxylase (AAB29520). The

CC invention also relates to the production of an L-amino acid via culture

CC of an organism of the invention. The method is useful for the production

CC of L-amino acids including L-lysine, L-valine, L-leucine, L-isoleucine

CC or L-threonine. The present sequence represents the Methylophilus

CC methylophilus diaminopimelate decarboxylase gene.

XX Sequence 2566 BP; 650 A; 668 C; 631 G; 616 T; 1 other:

XX Query Match 11.6%; Score 37.8; DB 21; Length 2566;

XX Best Local Similarity 58.4%; Pred. No. 0.038;

XX Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

XX 170 tcgaagaaggaggcaaacacacccctgagctcaaacctcagctgggagtgatg 229

XX 1082 tcaaaagccgcttgaacgagcattcttctcaacgtggaacacgtgaatgagctag 1141

XX 230 tgcacatcagatgactcgtgacgacgtccagaagaaggagcgcttccctc 282

XX 1142 accgcattcacaagtggtggcgccagcctgggcaaaaagcctatttccctc 1194

XX RESULT 8

AAT58627 standard; cDNA; 3984 BP.
ID AAT58627
XX
AC AAT58627;

DT 22-MAY-1997 (first entry)

XX Protein tyrosine phosphatase cDNA.

XX Protein tyrosine phosphatase; PTPH1; focal adhesion;

KM protein tyrosine kinase; malignancy; cancer; gene therapy;

XX retrovirus; vector; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 24..2765

FT /*tag= a

XX US5595911-A.

PD 21-JAN-1997.

PE 14-MAR-1990; 90US-0494036.

PR 01-MAR-1991; 91US-0663579.

PR 14-MAR-1990; 90US-0494036.

PR 16-AUG-1993; 93US-0107420.

XX (COLD-) COLD SPRING HARBOR LAB.

XX Tonks NK;

PI WPI: 1997-107583/10.

DR P-PSDB; AAM12522.

XX DNA encoding protein tyrosine phosphatase - for gene therapy of

PT cancer

XX Claim 1; Fig 1A-B; 12pp; English.

PS A cDNA clone (AAT58627) codes for a protein tyrosine phosphatase

CC (PTPH1) (AAM12522) which catalyses the dephosphorylation of proteins

CC in which tyrosyl residues have been phosphorylated through the

CC action of a protein tyrosine kinase (PTK). PTPH1 localises to

CC focal adhesions, a major site of action of oncogenic PTKs. The

CC cDNA clone was obtained by amplifying HeLa cDNA using primers based on

CC conserved segments within the catalytic domains of PTPHs, and using

CC a PCR product to screen an HeLa cDNA library. The PTPH1 cDNA can

CC be incorporated into a vector (pref. retroviral) to allow

CC expression of PTPH1 in mammalian cells in sufficient quantities to

CC overcome or counteract PTK activity. Phosphorylation of tyrosine

CC residues at abnormal levels is prevented or reversed, resulting in

CC the prevention or reversal of malignancy of cells.

XX Sequence 3984 BP; 1085 A; 954 C; 931 G; 1014 T; 0 other:

XX Query Match 11.6%; Score 37.8; DB 18; Length 3984;

XX Best Local Similarity 53.8%; Pred. No. 0.046;

XX Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

XX 97 gagggagagagcctccctgggttttactcctaaaggtgcttgagcagaggaaccatata 156

XX 1569 gatgaagatggaaaatttggatttaactcctaaaggagagtgatcaaaaagatgctctt 1628

XX 157 atcacctcctaaaggtcgaagaaggaggcaaacgacacccctgagctcacaactcgaagct 216

XX 1629 gtggtatacgaagataaacccagagtcacactcgtgagacacactgtatcccaagctgaagaa 1688

XX 217 ggggagtgaggttgacacatcaatg 241

XX 1689 ggggagtgaggttgacacatcaatg 1713

```

RESULT 9
AA78463
ID AAX78463 standard; cDNA; 3984 BP.
AC AAX78463;
XX
XX 07-SEP-1999 (first entry)
XX Human PTPH1 cDNA.
XX
XX PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;
KW localisation; treatment; overexpression; oncogenic; cell transformation;
KW prevention; phosphotyrosine; disease; malignant; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 24..2765
XX /*tag= a "PTPH1"
XX /product= "PTPH1"
XX /tag= b
XX /note= "Region of homology to the N-terminal domain
XX of band 4.1, ezrin and talin. This region is
XX known to be important for localisation to focal
XX adhesions"
XX
XX US5863781-A.
XX
XX 26-JAN-1999.
XX
XX 04-DEC-1996; 96US-0759536.
XX
XX 01-MAR-1991; 91US-0663579.
XX 14-MAR-1990; 90US-0494036.
XX 16-AUG-1993; 93US-0107420.
XX 04-DEC-1996; 96US-0759536.
XX
XX (COLD-) COLD SPRING HARBOR LAB.
XX
XX Tonks NK;
XX
XX WPI; 1999-131308/11.
XX P-PSDB; AAX25156.
XX
XX Protein tyrosine phosphatase PTPH1 - encoded by DNA of HeLa cells
XX
XX Disclosure; Fig 1A-B; 12pp; English.
XX
XX This sequence encodes a novel protein tyrosine phosphatase, PTPH1
XX isolated from HeLa cells. The protein of the invention appears to
XX localise to focal adhesions and is therefore potentially useful in the
XX treatment of cancer. Overexpression of PTPH1 can be used to counter the
XX effects of oncogenic protein tyrosine kinases such as those of
XX transforming viruses and for interfering with or reversing cell
XX transformation. This would provide a means of preventing or reversing
XX abnormally high levels of phosphotyrosine associated with any disease or
XX condition such as preventing or reversing malignancy associated with the
XX activity of a protein tyrosine kinase.
XX
XX Sequence 3984 BP; 1086 A; 954 C; 930 G; 1014 T; 0 other;
XX
XX Query Match 11.6%; Score 37.8; DB 20; Length 3984;
XX Best Local Similarity 53.8%; Pred. No. 0.046;
XX Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
XX
XX 97 gagggagagctccctgggttttacttaagggtggtcctgagcagcgagaaacatta 156
XX ||||| | | | | | | | | | | | | | | | | | | | | | |
XX 1569 gatgaagatggaaaattggatttaatttaaggagaggtggtatcaaaagatgctctt 1628
XX -

```

```

QY 157 atcatcttaaggtcgaagaaggggcaaaagcagacacacctgagctccaaactgcaggct 216
| | | | | | | | | | | | | | | | | | | | | |
Db 1629 gtgggtatcaaggataaaccacagagtcacctgcggacacctgattcttaagctgaacgaa 1688
| | | | | | | | | | | | | | | | | | | | | |
QY 217 ggggagtgaggttgcacatcaatg 241
| | | | | | | | | | | | | | | | | | | | | |
Db 1689 ggggatcaaatcgtgttaataatg 1713
| | | | | | | | | | | | | | | | | | | | | |

RESULT 10
AAC75454
ID AAC75454 standard; cDNA; 457 BP.
XX
XX AAC75454;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF1009 polynucleotide sequence SEQ ID NO:2017.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO2000058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX P-PSDB; AAB41245.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 5; Page 1531; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and aninaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:53:37 ; Search time 163.78 Seconds
(without alignments)
1711.719 Million cell updates/sec

Title: US-09-641-831-3

Perfect score: 327

Sequence: 1 atgataggaccactgaaga.....ggctgtagtgcgcagttga 327

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_1101.*

- 1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	327	100.0	327	22	Human CDNA encodin
2	323	98.8	402	22	Human CDNA encodin
3	323	98.8	573	22	Human CDNA encodin
4	64.6	19.8	462	22	Probe #5600 used t
5	55.8	17.1	114	22	Probe #19760 used
6	44.2	13.5	1525	21	Human pancreatic c
7	37.8	11.6	2566	21	Methylphilus meth
8	37.8	11.6	3984	18	Protein tyrosine p
9	37.8	11.5	3984	20	Human FRPHI CDNA
10	36.6	11.2	457	21	Human ORFX ORF1009
11	36.6	11.2	478	20	Human secreted pro

12	36.2	11.1	1696	20	AAX24307	Rat bone mineralis
13	36.2	11.1	1696	21	AAC87708	Rat LIM mineralisa
14	34.8	10.6	325	20	AAV88174	EST clone DU123..
15	34.6	10.6	717	20	AAX24320	Human bone mineral
16	34.6	10.6	717	21	AAC87713	Human LIM minerali
17	34.6	10.6	1456	21	AAC87741	Human LMP-2 (HLMP-
18	34.6	10.6	1575	21	AAC87742	Human LMP-3 (HLMP-
19	34.6	10.6	1620	21	AAC87737	Human LMP expressi
20	34.6	10.6	1644	21	AAX24322	Human bone mineral
21	34.6	10.6	1644	21	AAC87715	Human LIM minerali
22	34.6	10.6	1665	20	AAX24328	Human truncated bo
23	34.6	10.6	1665	21	AAC87738	Human LMP 5'-UTR s
24	34.6	10.6	1689	20	AAX24327	Human bone mineral
25	34.6	10.6	1689	21	AAC87727	Human LIM minerali
26	34	10.4	1511	19	AAV33198	Secreted protein D
27	33.8	10.3	1593	20	AAX242135	Human normal bladd
28	33.4	10.2	1597	17	AAT12908	Human insulin rece
29	33.2	10.2	1336	19	AAV30952	Homo sapiens HLIM-
30	33.2	10.2	1554	21	AAC75976	Human ORFX ORF1531
31	33.2	10.2	1606	21	AAF16133	Human prostate can
32	33.2	10.2	1844	22	AAH18032	Human cDNA sequenc
33	33.2	10.2	2987	9	AAH80456	Sequence of Rhodos
34	33	10.1	376	21	AAC02040	Human secreted pro
35	33	10.1	765	20	AAX24324	Human bone mineral
36	33	10.1	765	21	AAC87726	Human osteosarcoma
37	33	10.1	1619	17	AAT18796	Human immunophilin
38	33	10.1	3109	21	AAC78150	Human cancer assoc
39	32.6	10.0	1444	22	AAF23903	Human secreted pro
40	32.4	9.9	1960	16	AAQ84588	AMML chromosome in
41	32.4	9.9	2680	16	AAQ84590	AMML chromosome in
42	32.4	9.9	2887	16	AAQ84589	AMML chromosome in
43	32.4	9.9	6607	22	AAH98321	Rabbit EST-derived
44	32.4	9.9	31063	22	AAF28533	Genomic fragment #
45	32	9.8	2162	20	AAX40058	Colon cancer assoc

ALIGNMENTS

RESULT 1
ID AAS00040 standard; cDNA; 327 BP.
XX AAS00040;
AC AAS00040;
DT 11-MAY-2001 (first entry)
XX Human cDNA encoding novel human protein, NHP#2.
DE Human; novel human protein; NHP#2; gene therapy; drug screening;
KW obesity; high blood pressure; ss.
XX Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..327
FT /*tag= a
FT /product= "NHP#2"
XX WO200114422-A1.
XX 01-MAR-2001.
XX 18-AUG-2000; 2000WO-US22815.
XX Probe #5600 used t
XX 24-AUG-1999; 99US-0150511.
XX (LEXI-) LEXICON GENETICS INC.
XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
XX WPI; 2001-218430/22.
XX P-PSDB; AAU00032.

RESULT	13	
AAC87708		
ID	AAC87708 standard; cDNA; 1696 BP.	
XX		
XX	AAC87708;	
XX		
XX		
XX	01-MAR-2001 (first entry)	
XX		
DE	Rat LIM mineralisation protein (RLMP) encoding cDNA SEQ ID NO:2.	
XX		
XX	LIM mineralisation protein; LMP; bone formation; osteopathic;	
KW	osteogenic precursor cell; gene therapy; metabolic bone disease;	
KW	osteoporosis; bone degenerative disease; ss.	
XX		
XX	Rattus norvegicus.	
OS		
XX	WO200066178-A1.	
PN		
PD	09-NOV-2000.	
XX		
XX	28-APR-2000; 2000WO-US11664.	
PF		
XX		
XX	30-APR-1999; 99US-0132021.	
PR		
XX	{UYEM-} UNIV EMORY.	
XX		
XX	Boden SD, Hair GA;	
XX		
XX	WPI: 2000-672828/65.	
DR	P-PSDB; AAB36471.	
DR		
XX		
XX	New nucleic acid encoding a human LIM mineralization protein for	
PT	inducing or inhibiting bone formation, fusing a spine, stimulat	
PT		

RESULT	14
AAV88174	
ID	AAV88174 standard; cDNA; 325 BP.
XX	
AC	AAV88174;
XX	
DT	12-FEB-1999 (first entry)
XX	
DE	EST clone DU123.
XX	
KW	Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW	tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW	chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW	receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO9845437-A2.
XX	
PD	15-OCT-1998.
XX	
PF	10-APR-1998; 98WO-US06956.
XX	
PR	10-APR-1997; 97US-0837312.
XX	
PA	(GEMY) GENETICS INST INC.
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI	Racie LA, Spaulding V, Treacy M;
XX	

DR WPI; 1999-070078/06.

XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries

XX Claim 1; Page 308; 641pp; English.

CC The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haemostasis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.

XX Sequence 325 BP; 98 A; 71 C; 85 G; 71 T; 0 other;

Query Match 10.6%; Score 34.8; DB 20; Length 325;

Best Local Similarity 53.7%; Pred. No. 0.15; Mismatches 62; Indels 0; Gaps 0;

Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 55 aggcaccacgaaggaggtacattatctggagcattccctggagggagggcctccg 114

DB 2 aaggaacaacagggcagcagggatttctcgcattccatgcagcggggggcgcccg 61

QY 115 gtttactctaagggtgctgagcagcaggaaccattatcctaagggtcga 174

DB 62 gggttcagattgcaagggtgcaaggagcagcagcccttacaagttgcaagttcga 121

QY 175 gaagggggcaaacg 188

DB 122 aatcagagcaaacg 135

RESULT 15

AAAX24320

ID AAX24320 standard; cDNA; 717 BP.

XX AAX24320;

XX 07-JUN-1999 (first entry)

DE Human bone mineralisation protein LMP partial cDNA clone.

XX RAMP; LMP; LIM mineralisation protein; bone mineralisation; human;

KW cytokine; spine fusion; fracture repair; bone grafting;

KW osteoporosis; gene therapy; ss.

OS Homo sapiens.

XX WO9906563-A1.

XX 11-FEB-1999.

XX 29-JUL-1998; 98WO-US15814.

XX 02-APR-1998; 98US-0080407.

XX 30-JUL-1997; 97US-0054219.

XX (UYEM-) UNIV EMORY.

XX Boden S, Hair G;

XX WPI; 1999-153793/13.

XX LIM mineralisation proteins - used to induce bone formation from
PT osteogenic precursors, e.g. for treating osteoporosis or repairing
PT fractures

XX Example 17; Page 49; 67pp; English.

CC This is the nucleotide sequence of a human bone mineralisation
CC protein (LMP, or LIM mineralisation protein) partial cDNA clone
CC that was obtained from MG63 osteosarcoma cell cDNA using primers
CC (see AAX24317-18) respectively based on rat LMP cDNA (see AAX24307) and
CC on a previously isolated human partial clone (see AAX24319). A
CC full-length human LMP clone (see AAX24322) encoding a 457-amino acid
CC protein (see AAX27844) was subsequently obtained. LMP enhances bone
CC mineralisation in mammalian cells grown in vitro, and also induces
CC bone formation in vivo. Unlike other cytokines, LMP is not a
CC secreted protein, but is instead an intracellular signalling
CC molecule, so can provide intracellular signalling amplification and
CC easier assessment of transfected cells. Clinical applications
CC include enhancement of bone repair in fractures, bone defects, bone
CC grafting, and normal homeostasis in patients presenting with
CC osteoporosis. A method of inducing bone formation using
CC transfected osteogenic precursor cells is claimed.

XX Sequence 717 BP; 138 A; 262 C; 219 G; 98 T; 0 other;

Query Match 10.6%; Score 34.6; DB 20; Length 717;

Best Local Similarity 58.1%; Pred. No. 0.24; Mismatches 44; Indels 0; Gaps 0;

Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 86 aggcattctggagggagaggtccctgggttactctaagggtgcccggagcag 145

DB 14 aggtatgctgtagggcagcagcagccttgggtcctcaggtgcaaggggagacttca 73

QY 146 gagaacattatcatctctaagtcgaagaagggggcaaacag 190

DB 74 atgtgccctctccatttcccgctcactcctggggggcaaacgg 118

Search completed: January 31, 2002, 19:53:38
Job time: 191 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:50:03 : Search time 67.87 Seconds
(without alignments)
1091.179 Million cell updates/sec

Title: US-09-641-831-3

Perfect score: 327

Sequence: 1 atgctgagcaccactgaaga.....gctctgtatgcgcagtga 327

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued_Patents.NA.*

1: /cgn2.6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2.6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2.6/ptodata/2/1na/6A.COMB.seq:*
4: /cgn2.6/ptodata/2/1na/6B.COMB.seq:*
5: /cgn2.6/ptodata/2/1na/PCtus.COMB.seq:*
6: /cgn2.6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36.2	11.1	1696	4	US-09-124-238A-2
2	34.6	10.6	717	4	US-09-124-238A-7
3	34.6	10.6	1620	4	US-09-124-238A-32
4	34.6	10.6	1644	4	US-09-124-238A-9
5	34.6	10.6	1665	4	US-09-124-238A-33
6	34.6	10.6	1689	4	US-09-124-238A-22
7	34.6	10.2	1597	1	US-08-166-316-1
8	33.2	10.2	1336	2	US-08-739-485-6
9	33.2	10.1	765	4	US-09-124-238A-21
10	32.4	9.9	1960	2	US-08-533-306A-1
11	32.4	9.9	1960	2	US-08-742-923A-1
12	32.4	9.9	2680	2	US-08-533-306A-5
13	32.4	9.9	2680	2	US-08-742-923A-5
14	32.4	9.9	2887	2	US-08-533-306A-3
15	32.4	9.9	2887	2	US-08-742-923A-3
16	32.2	9.8	4403765	4	US-09-103-840A-2
17	32.2	9.8	4411529	4	US-09-103-840A-1
18	32.2	9.8	2162	3	US-08-948-705-5
19	32.2	9.8	2236	3	US-08-948-705-4
20	32.2	9.8	2289	3	US-08-948-705-8
21	31.8	9.7	1782	1	US-08-374-155A-13
22	31.8	9.7	1782	1	US-08-785-396-1
23	31.8	9.7	7218	1	US-08-232-463-14
24	30.8	9.4	1796	1	US-07-816-283-11
25	30.8	9.4	1796	1	US-08-417-103-11
26	29.6	9.1	1469	3	US-08-906-791-3
27	29.2	8.9	3924	3	US-08-726-214-9

28	29	8.9	3621	4	US-09-220-081-1	Sequence 1, Appl
29	28.6	8.7	4177	3	US-09-023-082A-23	Sequence 23, Appl
30	28.4	8.7	518	2	US-08-557-309B-1	Sequence 1, Appl
31	28.4	8.7	518	3	US-08-834-306-1	Sequence 1, Appl
32	28.4	8.7	518	4	US-08-993-674A-1	Sequence 1, Appl
33	28.4	8.7	1426	3	US-09-150-133-10	Sequence 10, Appl
34	28.4	8.7	1426	3	US-09-150-141-10	Sequence 10, Appl
35	28.4	8.7	1426	4	US-09-374-493-10	Sequence 10, Appl
36	28.4	8.7	1426	4	US-09-374-834-10	Sequence 10, Appl
37	28.4	8.7	1426	4	US-09-374-834-10	Sequence 10, Appl
38	28.4	8.7	2400	1	US-08-967-513-1	Sequence 1, Appl
39	28.4	8.7	2400	2	US-08-687-645B-1	Sequence 1, Appl
40	28.2	8.6	1255	2	US-08-766-551-6	Sequence 185, App
41	28	8.6	480	2	US-08-726-306A-185	Sequence 50, Appl
42	28	8.6	508	1	US-07-879-647A-50	Sequence 50, Appl
43	28	8.6	508	1	US-07-879-584A-50	Sequence 50, Appl
44	28	8.6	508	1	US-07-879-470A-50	Sequence 50, Appl
45	28	8.6	508	1	US-07-879-644A-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1
US-09-124-238A-2
Sequence 2, Application US/09124238A
Patent No. 6300127
GENERAL INFORMATION:
APPLICANT: Hair, Gregory A.
TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
FILE OF INVENTION: Expression Systems
FILE REFERENCE: 06148, 0115
CURRENT APPLICATION NUMBER: US/09/124,238A
CURRENT FILING DATE: 1998-07-29
PRIOR APPLICATION NUMBER: 60/054,219
PRIOR FILING DATE: 1997-07-30
PRIOR APPLICATION NUMBER: 60/080,407
PRIOR FILING DATE: 1998-04-02
NUMBER OF SEQ ID NOS: 36
SOFTWARE: MS Word
SEQ ID NO 2
LENGTH: 1696
TYPE: DNA
ORGANISM: Rattus norvegicus
US-09-124-238A-2
Query Match 11.1%; Score 36.2; DB 4; Length 1696;
Best Local Similarity 59.0%; Pred. No. 0.023; Mismatches 43; Indels 0; Gaps 0;
Matches 62; Conservative 0;
QY 86 aggcattctctgaggaagagcctcccttggtgttactctaaaggtgctgagcag 145
DB 86 aggtgctgtgaggaagcctcccttggtgttactctaaaggtgctgagcag 145
QY 146 gagaccattactctctctaaaggtcgaagagggcgaagcag 190
DB 146 acgtgccccctctctctctctgctcactcctctgaggaagcag 190
RESULT 2
US-09-124-238A-7
Sequence 7, Application US/09124238A
Patent No. 6300127
GENERAL INFORMATION:
APPLICANT: Hair, Gregory A.
TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
FILE OF INVENTION: Expression Systems
FILE REFERENCE: 06148, 0115
CURRENT APPLICATION NUMBER: US/09/124,238A
CURRENT FILING DATE: 1998-07-29


```

, PRIOR APPLICATION NUMBER: 60/054,219
, PRIOR FILING DATE: 1997-07-30
, PRIOR APPLICATION NUMBER: 60/080,407
, PRIOR FILING DATE: 1998-04-02
, NUMBER OF SEQ ID NOS: 36
, SOFTWARE: MS Word
, SEQ ID NO 7
, LENGTH: 717
, TYPE: DNA
, ORGANISM: Homo sapiens
US-09-124-238A-7

```

Query Match	10.6%	Score 34.6;	DB 4;	Length 717;
Best Local Similarity	58.1%;	Pred. No. 0.052;		
Matches 61;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;

QY	86	aggcattctcgaggggagctccctgggttttactctaagaagtgcgtgagcaacg	145
Db	14	aggtagtctgaggggcagcaccttggtcttcggctgcaaggggccaaggacttca	73
QY	146	gagaaaccattaatcatctcttaaggtcgaaaggggccaaaagcag	190
Db	74	atgtgccctctcatcttcccgctcactcctctggtggggcacaagcag	118

```

RESULT      3
US-09-124-238A-32
; Sequence 32, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Body
; TITLE OF INVENTION: Expression Syst
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SEQUENCE: MS Word
; SEQ ID NO 32
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-32

```

	Query Match	10.6%	Score 34.6;	DB 4;	Length 1620;
	Best Local Similarity	58.1%;	Pred. No. 0.075;		
	Matches	61;	Conservative	0;	Mismatches 44;
				Indels	0;
					Gaps
Qy	86	aggcaattcctggaggaggagctcctcgggggtttactctaaagggtggcctggagcaagc	145		
Db	14	aggtagtgcggaggccagcacctcggggcttcgcctgcgaaggggcgaagacttca	73		
Qy	146	gagaaaccattaatcatctctctaaagctcgaagaagggggcaagcag	190		
Db	74	atgtggccctctccattctccggctcactcctctgagagcaaacgag	118		

```

RESULT      4
US 09-124-238A-9
; Sequence 9, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
; TITLE OF INVENTION: Expression Systems

```

```

? FILE REFERENCE: 06148.0115
? CURRENT APPLICATION NUMBER: US/09/124,238A
? CURRENT FILING DATE: 1998-07-29
? PRIOR APPLICATION NUMBER: 60/054,219
? PRIOR FILING DATE: 1997-07-30
? PRIOR APPLICATION NUMBER: 60/080,407
? PRIOR FILING DATE: 1998-04-02
? NUMBER OF SEQ ID NOS: 36
? SOFTWARE: MS Word
? SEQ ID NO 9
? LENGTH: 1644
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-124-238A-9

```

	Query Match	10.6%	Score 34.5;	DB 4;	Length 1644;
	Best Local Similarity	56.1%	Pred. No. 0.076;		
	Matches 61;	Conservative	0;	Mismatches 44;	Indels 0; Gaps 0;
Qy	86	aggcattctcgaggagagctccctcggggtttactcaagggtggcctggagcag	145		
Db	14	aggtagtcgtggagggccacctcggggtctccgcgtgcaagggggcaagactcca	73		
Qy	146	gagcaaccatcaatcatctctaaagtcgaaggggggcaagcag	190		
Db	74	atgtgccccctctcatctccgcgtctactctgagggcaagcag	118		

```

RESULT      5
US-09-124-238A-33
; Sequence 33, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Scott A.
; APPLICANT: Boden, Gregg D.
; TITLE OF INVENTION: No. 6300127el Bo
; TITLE OF INVENTION: Expression Syst
; FILE REFERENCE: 0618.0115
; CURRENT APPLICATION NUMBER: US/09/12
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS Word
; SEQ ID NO 33
; LENGTH: 1665
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-33

```

[illegible]

RESULT 6
US-09-124-238A-22
; Sequence 22, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.

Ddb

69 AGGAAGCGGGCATGCCCCCAGACGGGTATCCGCCGGGCCTTGCAggc 128
||| | ||| ||| ||| ||| ||| ||| |||
QY

65 aggaagtlacatttactctgtagcaatccctgagagaaggatcccctgttgttacctc 124

QY 125 taaaggtgctgagcagcagagaccattatcatcttaaggtcgaagaaggggca 184
 Db 129 TCTCAGGGGCGATAGACTTCAACGACCTTTGGTCTACCCAGGATTACACAGGAACA 188
 QY 185 aagcag 190
 Db 189 AGCGG 194

RESULT 9
 US-09-124-238A-21
 ; Sequence 21, Application US/09124238A
 ; Patent No. 6300127
 ; GENERAL INFORMATION:
 ; APPLICANT: Hair, Gregory A.
 ; APPLICANT: Boden, Scott D.
 ; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
 ; TITLE OF INVENTION: Expression Systems
 ; FILE REFERENCE: 06148-0115
 ; CURRENT APPLICATION NUMBER: US/09/124,238A
 ; CURRENT FILING DATE: 1998-07-29
 ; PRIOR APPLICATION NUMBER: 60/054,219
 ; PRIOR FILING DATE: 1997-07-30
 ; PRIOR APPLICATION NUMBER: 60/080,407
 ; PRIOR FILING DATE: 1998-04-02
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: MS Word
 ; SEQ ID NO 21
 ; LENGTH: 765
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-124-238A-21

Query Match 10.1%; Score 33; DB 4; Length 765;
 Best Local Similarity 57.1%; Pred. No. 0.18;
 Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 86 aggcattctgagggagagctccctgttttactctaaaggtgctgagcag 145
 Db 74 aggtagtgctgagggcgagcacccttggttcgctgcaaggggcaaggacttca 133

QY 146 gagaccattatcatcttaaggtcgaagaaggggcaagcag 190
 Db 134 atgtgcctctccattctccggtccacctctggtgggcaagggcgcg 178

RESULT 10
 US-08-533-306A-1
 ; Sequence 1, Application US/08533306A
 ; Patent No. 5837457
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Pu
 ; APPLICANT: Collins, Francis S.
 ; APPLICANT: Siciliano, Michael J.
 ; APPLICANT: Claxton, David
 ; TITLE OF INVENTION: Markers for Detection of Chromosome 16
 ; TITLE OF INVENTION: Rearrangements
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/533,306A
 FILING DATE: September 25, 1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Smith, DeAnn F.
 REGISTRATION NUMBER: 36683
 REFERENCE/DOCKET NUMBER: 2115-00869COB
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (810) 641-1600
 TELEFAX: (810) 641-0270
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1960 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 INDIVIDUAL ISOLATE: Samples 3-6
 TISSUE TYPE: Acute myelomonocytic leukemia, M4eo
 TISSUE TYPE: subtype (inv16)
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT: 16(inv(16)(p13q22))
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1731
 PUBLICATION INFORMATION:
 AUTHORS: Liu, Pu
 AUTHORS: Tarle, Susan A.
 AUTHORS: Hajra, Amitav
 AUTHORS: Claxton, David F.
 AUTHORS: Mariton, Paula
 AUTHORS: Freedman, Matthew
 AUTHORS: Siciliano, Michael J.
 AUTHORS: Collins, Francis S.
 TITLE: Fusion between transcription factor
 TITLE: CBFb/PEBP2B and a myosin heavy chain in acute
 TITLE: myelomonocytic leukemia
 JOURNAL: Science
 DATE: August 1-1993
 RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 451 TO 534
 PUBLICATION INFORMATION:
 AUTHORS: Liu, Pu
 AUTHORS: Claxton, David F.
 AUTHORS: Mariton, Paula
 AUTHORS: Hajra, Jeanette
 AUTHORS: Freedman, Matthew
 AUTHORS: Chandrasekharappa, Settara C.
 AUTHORS: Yanagisawa, Kohsuke
 AUTHORS: Stallings, Raymond L.
 AUTHORS: Collins, Francis S.
 AUTHORS: Siciliano, Michael J.
 TITLE: Identification of yeast artificial
 TITLE: chromosomes containing the inversion 16 p-arm
 TITLE: breakpoint associated with acute myelomonocytic
 TITLE: leukemia
 JOURNAL: Blood
 DATE: 1993
 US-08-533-306A-1

Query Match 9.9%; Score 32.4; DB 2; Length 1960;
 Best Local Similarity 48.9%; Pred. No. 0.43;
 Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 124 ctaaaggtgctgagcagcagagaccattatcatcttaaggtcgaagaaggggca 183
 Db 778 CTGGAAGGGGACCTGAAAGACCTTGAGCTTCAGGGCGACTCTGCCATCAAGGGGAG 837
 QY 184 aaagcagacccctgagctcccaactgcaggtggtgggtgaggtgtgtgcacatcaatgag 243


```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Sample 2
; TISSUE TYPE: Acute myelomonocytic leukemia, M4EO
; TISSUE TYPE: subtype (inv16)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 16[inv(16)(p13q22)]
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2451
; US-08-533-306A-5

Query Match 9.9%; Score 32.4; DB 2; Length 2680;
Best Local Similarity 48.9%; Pred. No. 0.49;
Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps

Qy 124 ctaaaggggtggccttgagcagcgagaaacattaatcattctctaaggttcgaagaagggggc 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1498 CTGAGAGGGGACCTGAAGACCTTGAGCTTCAGGCCGACTCTGCCATCAAGGGAGGGAG 1557
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 184 aaagcagacacctgagctccaaactgcaggctggggatgaggttgccacatcaatgag 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1558 GAAGCCATCAAGCAGCTAGCGCAACTGCAGGCTCAGATGAAGGACTTTTCAAGAGAGCTG 1617
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 244 gtgaactctagcagctccagaagaagcagcagtttccctcgttgaaggatcctacaaga 301
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1618 GAAGATGCCCGTGCCTCCAGATGATGATCTTTGCCACAGCCCAAGCAATGAGAAGA 1675
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-08-742-923A-5
; Sequence 5, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: No. 5869611ember 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683

```

```

; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Sample 2
; TISSUE TYPE: Acute myelomonocytic leukemia, M4Eo
; TISSUE TYPE: subtype (inv16)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 16(inv(16)(p13q22))
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2451
; US-08-742-923A-5

Query Match          9.9%; Score 32.4; DB 2; Length 2680;
Best Local Similarity 48.9%; Pred. No. 0.49;
Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 124 ctaaagggtggcctggagcacggagaccattaatatctcttaagggtcgagaagggggc 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1498 CTGAAGGGGACCTGGAAGACCTGGAGCTTTCAGCCGACTCTGCCATCAAGGGGAGGGAG 1557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 aaagcagacacccctgagctccaaactcaggctgggggatgaggtgtgcacatcaatgag 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1558 GAACCCATCAGCAGCTACGCAAACTCGAGCTCAGATGAAGGACTTTCAAAGNAGACTG 1617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 gtgactctgagcagctccagaaaggaggcaggttccctgtgtgaaggatcctcaaga 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1618 GAAGATCCCCGTGCTCCAGAGATGAGATCTTTGCCACAGCCAAAGAGAATGAGAAGA 1675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14
 US-08-533-306A-3
 ; Sequence 3, Application US/08533306A
 ; Patent No. 5837457
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Pu
 ; APPLICANT: Collins, Francis S.
 ; APPLICANT: Siciliano, Michael J.
 ; APPLICANT: Claxton, David
 ; TITLE OF INVENTION: Markers for Detection of Chromosome 16
 ; TITLE OF INVENTION: Rearrangements
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/533.306A
 ; FILING DATE: September 25, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Deann F.

REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869COB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2887 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE: Sample 1
TISSUE TYPE: Acute myelomonocytic leukemia, M4Eo
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 16(inv(16)(p13q22))
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2658
US-08-533-306A-3

Query Match 9.9%; Score 32.4; DB 2; Length 2887;
Best Local Similarity 48.9%; Pred. No. 0.51; Mismatches 87; Conservative 0; Indels 91; Gaps 0;

124 ctaagggtgctgagcagcaggaaccattatctcttaagtcgaagaggggc 183
1705 CTGGAAGGGGACCTGAAAGACCTGAGCTTCAGCCGACCTGCAATCAAGGGAG 1764
184 aaagcagaccctgagctccaaactcgaagctgagtgatgagtgatgacatcatgag 243
1765 GAAGCCATCAAGCAGCTACGCAAACTGACGCTCAGATGAAGGACTTTCAAGAGAGCTG 1824
244 gtgactctgagcagctccagaagaagagcagttccctggtgaaagatcctacaaga 301
1825 GAAGTGCCTGCTCCAGAGATGAGATCTTTGCGCACAGCCAAAGAAATGAGAGA 1882

RESULT 15

US-08-742-923A-3

Sequence 3, Application US/08742923A

Patent No. 5869611

GENERAL INFORMATION:

APPLICANT: Liu, Pu

APPLICANT: Collins, Francis S.

APPLICANT: Siciliano, Michael J.

TITLE OF INVENTION: Markers for Detection of Chromosome 16

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESS: Harness, Dickey & Pierce, P.L.C.

STREET: P.O. Box 828

CITY: Bloomfield Hills

STATE: MI

COUNTRY: USA

ZIP: 48303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

APPLICATION NUMBER: US/08/742,923A

FILING DATE: No. 5869611ember 1, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869DVC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2887 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE: Sample 1
TISSUE TYPE: Acute myelomonocytic leukemia, M4Eo
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 16(inv(16)(p13q22))
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2658
US-08-742-923A-3

Query Match 9.9%; Score 32.4; DB 2; Length 2887;
Best Local Similarity 48.9%; Pred. No. 0.51; Mismatches 87; Conservative 0; Indels 91; Gaps 0;

124 ctaagggtgctgagcagcaggaaccattatctcttaagtcgaagaggggc 183
1705 CTGGAAGGGGACCTGAAAGACCTGAGCTTCAGCCGACCTGCAATCAAGGGAG 1764
184 aaagcagaccctgagctccaaactcgaagctgagtgatgagtgatgacatcatgag 243
1765 GAAGCCATCAAGCAGCTACGCAAACTGACGCTCAGATGAAGGACTTTCAAGAGAGCTG 1824
244 gtgactctgagcagctccagaagaagagcagttccctggtgaaagatcctacaaga 301
1825 GAAGTGCCTGCTCCAGAGATGAGATCTTTGCGCACAGCCAAAGAAATGAGAGA 1882

Search completed: January 31, 2002, 19:50:07
Job time: 100 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:53:38 ; Search time 163.78 Seconds

(Without alignments)
2104.315 Million cell updates/sec

Title: US-09-641-831-5

Perfect score: 402
Sequence: 1 atgatgagaccactgaaga.....atcatgcactgcagccttga 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	402	100.0	402	22	AA500041 Human cDNA encoding
2	323	80.3	327	22	AA500040 Human cDNA encoding
3	323	80.3	573	22	AA500039 Human cDNA encoding
4	64.6	16.1	462	22	AA137914 Probe #6600 used t
5	59.2	14.7	32134	22	AA163522 Human kidney relat
6	59.2	14.7	32134	22	AA163523 Human kidney relat
7	56.2	14.0	634	22	AA163522 Human kidney relat
8	55.8	13.9	114	22	AA151074 Probe #19760 used
9	53.8	13.4	17131	21	AAZ60888 DNA encoding a hum
10	52.8	13.1	391	21	AAZ17932 Human secreted pro
11	52.4	13.0	50000	21	AA596367 Polymorphic repeat

C 12	52	12.9	608	20	AA520431 Human secreted pro
C 13	51.6	12.8	1159	22	AA594464 Human foetal cDNA,
C 14	50.6	12.6	313	16	AA23935 Human gene signatu
C 15	50.6	12.6	362	22	AA111740 Probe #1673 for ge
C 16	50.6	12.6	362	22	AA133047 Probe #1733 used t
C 17	50.6	12.6	362	22	AA101668 Probe #1659 used t
C 18	49.2	12.2	385	21	AA16802 Human secreted pro
C 19	49.2	12.2	1314	21	AA579722 Human secreted pro
C 20	49.2	12.2	72604	20	AA10752 Genomic sequence o
C 21	49	12.2	335	21	AA30196 Human secreted pro
C 22	48.4	12.0	1659	18	AA18444 Human cDNA sequenc
C 23	48.4	12.0	87350	22	AA583003 Human WRN genomic
C 24	48	11.9	223	21	AA24831 Human secreted pro
C 25	47	11.7	406	21	AA26743 Human secreted pro
C 26	47	11.7	8353	18	AA26936 Methods for diagno
C 27	47	11.7	44453	20	AA23519 Human kidney amino
C 28	46.8	11.6	3205	22	AA173404 Human R33 protein
C 29	46.6	11.6	28994	19	AA15826 Genomic DNA for in
C 30	46.6	11.6	33780	22	AA24652 Human EST-derived
C 31	46.2	11.5	1691	22	AA59051 Human neuroblastom
C 32	46.2	11.5	1691	22	AA59080 Human EST-derived
C 33	45.8	11.4	3530	22	AA57847 I-FLICE-2 coding s
C 34	45	11.2	2597	19	AA54807 Human foetal cDNA,
C 35	44.8	11.1	488	22	AA54277 Human secreted pro
C 36	44.6	11.1	172	21	AA16201 Human secreted pro
C 37	44.4	11.0	231	21	AA19578 3'-flanking sequen
C 38	44.4	11.0	472	17	AA12389 Human CTLR-L DNA,
C 39	44.4	11.0	2143	19	AA61937 Human FLICE-like 1
C 40	44.4	11.0	2143	20	AA239040 Human FLN-1 encod1
C 41	44.4	11.0	2188	19	AA521375 Human G1 protein 1
C 42	44.4	11.0	2243	19	AA52968 Human cDNA sequenc
C 43	44.4	11.0	3530	22	AA18408 Human purh gene ge
C 44	44.4	11.0	41684	21	AA28150 Human secreted pro
C 45	44.2	11.0	1089	20	AA37460 Human secreted pro

ALIGNMENTS

RESULT 1	AA500041	standard; cDNA; 402 BP.
ID	AA500041	standard; cDNA; 402 BP.
NC	AA500041;	
DT	11-MAY-2001	(first entry)
XX		
DE	Human cDNA encoding novel human protein, NHP#3.	
XX		
KW	Human; novel human protein; NHP#3; gene therapy; drug screening;	
KW	obesity; high blood pressure; ss.	
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..402
FT		/*tag= a
FT		/product= "NHP#3"
PN	WO200114422-A1.	
XX		
PD	01-MAR-2001.	
XX		
PF	18-AUG-2000; 2000WO-US22815.	
XX		
PR	24-AUG-1999; 99US-0150511.	
XX		
PA	(LEXI-) LEXICON GENETICS INC.	
XX		
PI	Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;	
XX		
DR	WPI; 2001-218430/22.	
DR	P-PSDB; AAU00033.	

XX Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX
 XX
 XX Claim 5; Page 25; 29pp; English.
 XX
 XX The sequence encodes a novel human protein, NHP#3, which shares
 CC structural motifs with human APXL protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide probes can be used to
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 XX
 XX Sequence 402 BP; 110 A; 91 C; 117 G; 84 T; 0 other;

Query Match 100.0%; Score 402; DB 22; Length 402;
 Best Local Similarity 100.0%; Pred. No. 6.3e-120;
 Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgagagacacactgaagactccacagcctagtgccacattaaacttaacacggcc 60
 Db 1 atgagagacacactgaagactccacagcctagtgccacattaaacttaacacggcc 60
 QY 61 accaagggaaggtacatttctctgagggcattctgagggagagctccctgggtttt 120
 Db 61 accaagggaaggtacatttctctgagggcattctgagggagagctccctgggtttt 120
 QY 121 actctaaagggtgctgagacgagcaggaacattatctctctgaagaagg 180
 Db 121 actctaaagggtgctgagacgagcaggaacattatctctctgaagaagg 180
 QY 181 ggcaagacagacacccctgagctccaaactgcagctgggagtggtgtgacatcaat 240
 Db 181 ggcaagacagacacccctgagctccaaactgcagctgggagtggtgtgacatcaat 240
 QY 241 gagtgactctgagcgtccgaaaggagggcagtttccctgggtgaaaggatctacaag 300
 Db 241 gagtgactctgagcgtccgaaaggagggcagtttccctgggtgaaaggatctacaag 300
 QY 301 accctcaggtgtagtgcgagaaatgggttctgtctgttgcgcagaatgggaaggta 360
 Db 301 accctcaggtgtagtgcgagaaatgggttctgtctgttgcgcagaatgggaaggta 360
 QY 361 gtggtattcatagcagcatcatcatcagctgcagccttga 402
 Db 361 gtggtattcatagcagcatcatcatcagctgcagccttga 402

RESULT 2
 AAS00040
 ID AAS00040 standard; cDNA; 327 BP.
 XX
 AC AAS00040;
 XX
 DT 11-MAY-2001 (first entry)
 XX
 DE Human cDNA encoding novel human protein, NHP#2.
 XX
 KW Human; novel human protein; NHP#2; gene therapy; drug screening;
 KW obesity; high blood pressure; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT CDS 1..327

FT
 FT /*tag= a
 XX /product= "NHP#2"
 PN WO200114422-A1.
 XX
 XX 01-MAR-2001.
 XX
 XX 18-AUG-2000; 2000WO-US22815.
 XX
 XX 24-AUG-1999; 99US-0150511.
 XX (LEXI-) LEXICON GENETICS INC.
 XX
 XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
 XX
 XX WPI; 2001-218430/22.
 XX P-FSDB; AAO00032.
 XX
 XX Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX
 XX Claim 3; Page 25; 29pp; English.
 XX
 XX The sequence encodes a novel human protein, NHP#2, which shares
 CC structural motifs with human APXL protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide probes can be used to
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 XX
 XX Sequence 327 BP; 91 A; 76 G; 96 C; 64 T; 0 other;

Query Match 80.3%; Score 323; DB 22; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.8e-94;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgagagacacactgaagactccacagcctagtgccacattaaacttaacacggcc 60
 Db 1 atgagagacacactgaagactccacagcctagtgccacattaaacttaacacggcc 60
 QY 61 accaagggaaggtacatttctctgagggcattctgagggagagctccctgggtttt 120
 Db 61 accaagggaaggtacatttctctgagggcattctgagggagagctccctgggtttt 120
 QY 121 actctaaagggtgctgagacgagcaggaacattatctctctgaagaagg 180
 Db 121 actctaaagggtgctgagacgagcaggaacattatctctctgaagaagg 180
 QY 181 ggcaagacagacacccctgagctccaaactgcagctgggagtggtgtgacatcaat 240
 Db 181 ggcaagacagacacccctgagctccaaactgcagctgggagtggtgtgacatcaat 240
 QY 241 gagtgactctgagcgtccgaaaggagggcagtttccctgggtgaaaggatctacaag 300
 Db 241 gagtgactctgagcgtccgaaaggagggcagtttccctgggtgaaaggatctacaag 300
 QY 301 accctcaggtgtagtgcgag 323
 Db 301 accctcaggtgtagtgcgag 323

RESULT 3
 AAS00039
 ID AAS00039 standard; cDNA; 573 BP.
 XX

RESULT 5
AAI63522/c
ID AAI63522 standard; DNA; 32134 BP.
XX
AC AAI63522;
XX 22-OCT-2001 (first entry)
XX Human kidney related polynucleotide SEQ ID NO 837.
XX
KW Human: kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neotropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; vulnery; anticonvulsant; antiparasitic;
KW gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; ds.
XX
XX Homo sapiens.
XX
XX WO20015323-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01343.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184666.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225759.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 06-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.

17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250161.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-488784/53.
XX
XX New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
XX Disclosure; SEQ ID NO 837; 564bp + Sequence Listing; English.
XX
XX The invention relates to novel kidney related polynucleotides
CC (AA163523/3) and the encoded polypeptides (AA42417-AA42691)
CC collectively known as kidney antigens and the use of such kidney antigens
CC for detecting disorders of the kidney, especially kidney cancer and
CC kidney cancer metastases. The polynucleotides and proteins are also
CC useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. The genes are isolated from a range
CC of human tissues disclosed in the specification. The nucleic acids,
CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
CC and other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 32134 BP; 8893 A; 6674 C; 6825 G; 9742 T; 0 other;
SQ
Query Match 14.7%; Score 59.2; DB 22; Length 32134;
Best Local Similarity 83.8%; Pred. No. 1.9e-08;
Matches 67; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY 323 gaatgggggtctgtgctatgtccagaatgaagtaagtgctatctatagcatgac 382
DB 22933 GAGATAGGGTCTTGTCTATTTTCCAGCGCGGATTCGACGTGCTATTCATGAGCATGAG 22873
OY 383 atcatgcactgcagccttga 402
DB 22872 ATAGTGCACCTGTAACCTTGA 22853
RESULT 6
ID AA163523/C
XX AA163523 standard; DNA; 32192 BP.

AC AA163523;
XX
DT 22-OCT-2001 (first entry)
DE Human kidney related polynucleotide SEQ ID NO 838.
XX
XX Human: kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neotropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; anticonvulsant; antiparasitic;
KW gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; ds.
XX
OS Homo sapiens.
XX
PN W020015323-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01343.
PF
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0228343.
PR 01-SEP-2000; 2000US-0228344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.

XX Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
 KM ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
 KM Insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
 KM Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
 KM thyroiditis; postpartum thyroiditis; rheumatoid arthritis;
 KM Hashimoto's disease; coeliac disease; ss.
 XX Homo sapiens.
 OS
 PN WO200056856-A2.
 PD 28-SEP-2000.
 PF 24-MAR-2000; 2000MO-US07938.
 XX
 PR 25-MAR-1999; 99US-0126215.
 XX
 PA (GEMT) GENETICS INST INC.
 XX
 PI Ling V, Wu P, Gray GS;
 XX
 DR WPI; 2000-628257/60.
 XX
 PT Determining predisposition of humans to develop autoimmune disease
 PT Involves detecting polymorphic microsatellite repeat sequence within
 PT human costimulatory receptor gene locus
 XX
 PS Disclosure; Page 128-142; 160pp; English.
 XX
 CC Two human bacterial artificial chromosome (BAC) clones that included
 CC and flanked the human CTLA-4 locus were cloned and sequenced. The
 CC sequence data was assembled into a contiguous sequence that is presented
 CC in AA96363-68. AA96363-64 comprise BAC clone 22700, and AA96365-68
 CC comprise BAC clone 22608. The sequences contain polymorphic
 CC microsatellite repeat (PMR) sequences. The specification describes a
 CC method for determining the predisposition of a human subject to develop
 CC autoimmune disease. The method comprises detecting a PMR sequence in the
 CC CD28, ICOS gene or CTLA4 gene of the human costimulatory receptor gene
 CC locus (hcgRL). PMR sequences vary in length among individuals and can be
 CC amplified to generate products that differ in size. These products can
 CC then be detected by rapid and convenient high resolution processes. The
 CC method is useful for determining the predisposition of insulin-dependent
 CC diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune
 CC hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
 CC postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,
 CC coeliac disease and leprosy. PMR sequences within hcgRL are useful as
 CC markers in a variety of assays and in the field of forensic medicine,
 CC disease diagnosis and human genome mapping.
 CC
 XX
 SQ Sequence 50000 BP; 14612 A; 9948 C; 10072 G; 15368 T; 0 other;

XX
 DE Human secreted protein gene 20.
 XX
 KM Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
 KM developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KM* inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
 KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN WO9906423-A1.
 PD 11-FEB-1999.
 PF 29-JUL-1998; 98MO-US15949.
 XX
 PR 19-AUG-1997; 97US-0056730.
 PR 30-JUL-1997; 97US-0054209.
 PR 30-JUL-1997; 97US-0054211.
 PR 30-JUL-1997; 97US-0054212.
 PR 30-JUL-1997; 97US-0054213.
 PR 30-JUL-1997; 97US-0054214.
 PR 30-JUL-1997; 97US-0054215.
 PR 30-JUL-1997; 97US-0054217.
 PR 30-JUL-1997; 97US-0054218.
 PR 30-JUL-1997; 97US-0054234.
 PR 30-JUL-1997; 97US-0054236.
 PR 18-AUG-1997; 97US-0055968.
 PR 18-AUG-1997; 97US-0055969.
 PR 18-AUG-1997; 97US-0055972.
 PR 19-AUG-1997; 97US-0056534.
 PR 19-AUG-1997; 97US-0056543.
 PR 19-AUG-1997; 97US-0056554.
 PR 19-AUG-1997; 97US-0056561.
 PR 19-AUG-1997; 97US-0056727.
 PR 19-AUG-1997; 97US-0056729.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Carter KC, Endress GA, Fan P, Feng P, Kyaw H, Lafleur DW;
 PI Li Y, Moore PA, Rosen CA, Ruben SM, Shi Y, Wei Y;
 PI Zeng Z;
 XX
 DR WPI; 1999-153691/13.
 DR P-PSDB; AAY00277.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1; Page 219; 312pp; English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number is given in the descriptor line.
 CC The gene can be used to generate fusion proteins by linking to the gene
 CC to a human immunoglobulin Fc portion (e.g. AAX20403) for increasing the
 CC stability of the fused protein as compared to the human protein only.
 CC The invention relates to 83 novel genes and their fragments (nucleic
 CC acid sequences: AAX20412-X20499; amino acid sequences: AAY00258-Y00377)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 86
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX20412 for described uses).
 CC
 XX
 SQ Sequence 608 BP; 162 A; 192 C; 113 G; 138 T; 3 other;

Matches	02	Conservative	0	Mismatches	15	Indels	0	Stop
Qy	322	agaaatggggtctgtctatgtgccagaatggaaggtagtgctattcattaggcattgat	381					

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:48:27 ; Search time 67.87 seconds
(without alignments)
1912.066 Million cell updates/sec

Title: US-09-641-831-1
573
Sequence: 1 atgatggagccactgaaga.....gagcctatttaagtagatga 573

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 11323899 residues
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/prodata/2/1na/5A_COMB.seq:*
2: /cgn2_6/prodata/2/1na/5B_COMB.seq:*
3: /cgn2_6/prodata/2/1na/6A_COMB.seq:*
4: /cgn2_6/prodata/2/1na/6B_COMB.seq:*
5: /cgn2_6/prodata/2/1na/PCOTUS_COMB.seq:*
6: /cgn2_6/prodata/2/1na/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37.4	6.5	717	4 US-09-124-238A-7	Sequence 7, Appl
2	37.4	6.5	1620	4 US-09-124-238A-32	Sequence 32, Appl
3	37.4	6.5	1644	4 US-09-124-238A-9	Sequence 9, Appl
4	37.4	6.5	1665	4 US-09-124-238A-33	Sequence 33, Appl
5	37.4	6.5	1689	4 US-09-124-238A-22	Sequence 22, Appl
6	36.2	6.3	1696	4 US-09-124-238A-2	Sequence 2, Appl
7	35.8	6.2	7218	1 US-08-232-463-14	Sequence 14, Appl
8	34.2	6.0	765	4 US-09-124-238A-21	Sequence 21, Appl
9	33.8	5.9	2400	1 US-08-967-513-1	Sequence 1, Appl
10	33.8	5.9	2400	2 US-08-687-645B-1	Sequence 1, Appl
11	33.4	5.8	1597	1 US-08-166-316-1	Sequence 1, Appl
12	33.4	5.8	2162	3 US-08-948-705-5	Sequence 5, Appl
13	33.4	5.8	2236	3 US-08-948-705-4	Sequence 4, Appl
14	33.4	5.8	2289	3 US-08-948-705-8	Sequence 8, Appl
15	33.2	5.8	1336	2 US-08-739-485-6	Sequence 6, Appl
16	32.4	5.7	1960	2 US-08-533-306A-1	Sequence 1, Appl
17	32.4	5.7	1960	2 US-08-742-923A-1	Sequence 5, Appl
18	32.4	5.7	2680	2 US-08-533-306A-5	Sequence 5, Appl
19	32.4	5.7	2680	2 US-08-742-923A-5	Sequence 5, Appl
20	32.4	5.7	2887	2 US-08-533-306A-3	Sequence 3, Appl
21	32.4	5.7	2887	2 US-08-742-923A-3	Sequence 3, Appl
22	32.2	5.6	4403765	4 US-09-103-840A-2	Sequence 2, Appl
23	32.2	5.6	4411529	4 US-09-103-840A-1	Sequence 1, Appl
24	31.8	5.5	1782	1 US-08-374-155A-13	Sequence 13, Appl
25	31.8	5.5	1782	2 US-08-785-336-13	Sequence 13, Appl
26	30.8	5.4	1796	1 US-07-816-283-11	Sequence 11, Appl
27	30.8	5.4	1796	1 US-08-417-103-11	Sequence 11, Appl

C	28	30.8	5.4	3924	3 US-08-726-214-9	Sequence 9, Appl
C	29	30.6	5.3	2008	4 US-08-637-823B-3	Sequence 3, Appl
C	30	30.2	5.3	450	3 US-08-586-039B-46	Sequence 46, Appl
C	31	30.2	5.3	465	3 US-08-586-039B-40	Sequence 40, Appl
C	32	30.2	5.3	513	3 US-08-586-039B-44	Sequence 44, Appl
C	33	30.2	5.3	1645	2 US-08-039-297B-1	Sequence 1, Appl
C	34	30	5.2	508	1 US-07-879-647A-30	Sequence 50, Appl
C	35	30	5.2	508	1 US-07-879-584A-50	Sequence 50, Appl
C	36	30	5.2	508	1 US-07-879-470A-50	Sequence 50, Appl
C	37	30	5.2	508	1 US-07-879-644A-50	Sequence 50, Appl
C	38	30	5.2	508	1 US-07-879-640A-50	Sequence 50, Appl
C	39	30	5.2	508	1 US-07-879-594A-50	Sequence 50, Appl
C	40	30	5.2	508	1 US-07-879-469A-25	Sequence 50, Appl
C	41	30	5.2	1744	1 US-07-879-647A-25	Sequence 25, Appl
C	42	30	5.2	1744	1 US-07-879-584A-25	Sequence 25, Appl
C	43	30	5.2	1744	1 US-07-879-470A-25	Sequence 25, Appl
C	44	30	5.2	1744	1 US-07-879-644A-25	Sequence 25, Appl
C	45	30	5.2	1744	1 US-07-879-640A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-124-238A-7
Sequence 7, Application US/09124238A
Patent No. 6300127
GENERAL INFORMATION:
APPLICANT: Hair, Gregory A.
TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
FILE REFERENCE: 06148.0115
CURRENT APPLICATION NUMBER: US/09/124,238A
PRIOR FILING DATE: 1998-07-29
PRIOR APPLICATION NUMBER: 60/054,219
PRIOR FILING DATE: 1997-07-30
PRIOR APPLICATION NUMBER: 60/080,407
NUMBER OF SEQ ID NOS: 36
SOFTWARE: MS Word
SEQ ID NO 7
LENGTH: 717
TYPE: DNA
ORGANISM: Homo sapiens
US-09-124-238A-7

Query Match 6.5% Score 37.4; DB 4; Length 717;
Best Local Similarity 49.8% Pred. No. 0.01; Indels 6; Gaps 1;
Matches 126; Conservative 0; Mismatches 121;
Query 86 aggcattcttgaggagagagcctcccttggtttactctaaggttgctgagcag 145
Db 14 aggttgcttgaggagagagcctcccttggtttactctaaggttgctgagcag 73
Query 146 gagaacccaatcatctcctaagtcgaagaaggagcagaacacccctgagctca 205
Db 74 atgtgccctccatctccctccgctcactcccttggttggaagcgcgc-----gacagcgc 127
Query 206 aactcagagccttggttgatgtgtgacatcaatgaagtgactcttgagcagctcagaa 265
Db 128 gatttgccgttggttgactggttgctgagcagcagtcgaggaatgctggtgagctcacaac 187
Query 266 aggagagcagttcccttggttgaaagatcctaagagccctcagctggtgagcagc 325
Db 188 aactcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 247
Query 326 tctcccccagcgt 338
Db 248 gggcccccagcgt 260

```

RESULT      2
US-09-124-238A-32
; Sequence 32, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
; TITLE OF INVENTION: Expression Systems
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS word
; SEQ ID NO 32
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-32

```

Query Match	6.5%	Score 37.4	DB 4	Length 1620
Best Local Similarity	49.8%	Pred. No. 0.016		
Matches 126	Conservative	0	Mismatches 121	Indels 6
Gaps				
Qy	86	aggatttcctggaggagagctccctgggttttactctaaagggtggcctgagacag	145	
Db	14	aggtatgtctggaggggccacacttggggcttccggctcaagggggcaagacttca	73	
Qy	146	gagaaccattaatctctctaaggtctgaagaagggggcaaacgacacacctgagctcca	205	
Db	74	atgtgccctctccattctccggctcactctctggggcgaagcggc-----cgagcgg	127	
Qy	206	aactgcagggttggggatgaggttgtcacatcaatgagtgactctgagcagctccagaa	265	
Db	128	gagtgggcgtgggtgactgggtgctgagcatcgatgcgagaaTgcgggtagcctcac	187	
Qy	266	aggaggcagtttccctggtgaaggatctctacaagacctcaggctggtagtgcgcgac	325	
Db	188	accatgaagctcgaacaagaatccgggctcggggagcgctcagcctggggctcagca	247	
Qy	326	tctccccaccggt	338	
Db	248	gggcccaagccgt	260	

RESULT 3
 US-09-124-238A-9
 ; Sequence 9, Application US/09124238A
 ; Patent No. 6300127
 ; GENERAL INFORMATION:
 ; APPLICANT: Hair, Gregory A.
 ; APPLICANT: Boden, Scott D.
 ; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
 ; TITLE OF INVENTION: Expression Systems
 ; FILE REFERENCE: 06148.0115
 ; CURRENT APPLICATION NUMBER: US/09/124,238A
 ; CURRENT FILING DATE: 1998-07-29
 ; PRIOR APPLICATION NUMBER: 60/054,219
 ; PRIOR FILING DATE: 1997-07-30
 ; PRIOR APPLICATION NUMBER: 60/080,407
 ; PRIOR FILING DATE: 1998-04-02
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: MS Word
 ; SEQ ID NO 9
 ; LENGTH: 1644
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-124-238A-9

	Query Match	6.5%;	Score 37.4;	DB 4;	Length 1644;
	Best Local Similarity	49.8%;	Pred. No. 0, 016;		
	Matches 126;	Conservative	0;	Mismatches 121;	Indels 6; Gaps 1;
QY	86	aggcattctcgaggagagtcctccctgggttttactctaagaagtgctgagcacg	145		
Dd	14	aggtagtctgaggggcacaccttgggtccgctgcaagggggcaaggaattcca	73		
QY	146	gagaaaccataatcatcttaaggtcgagaagggggcaagcgacacccctgagctoca	205		
Dd	74	argtgccccctccattcccoggtccactctgggggcaagcggc-----gcaggcg	127		
QY	206	aactgcaggctgggatgaggtgtgacatcatactgagtgactctgagcagctccagaa	265		
Dd	128	gagtgcccgtaggtgaactgggtctgagcatcgatgccggaatgccggtagcctcac	187		
QY	266	aggagggcagtttccctggtgaaggatcctacaagaccctcaggctggtagtcgcagcc	325		
Dd	188	acatcgagctcagaacaagatccggcctcggggagcgccctcagcctgggctcagca	247		
QY	326	tctcccaaccggt	338		
Dd	248	gggccagccggt	260		

RESULT 4
US-09-124-238A-33
; Sequence 33, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
; TITLE OF INVENTION: Expression Systems
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS Word
; SEQ ID NO 33
; LENGTH: 1665
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-33

Query Match	6.5%	Score 37.4;	DB 4;	Length 1665;
Best Local Similarity	49.8%;	Pred. No. 0.016;		
Matches 126;	Conservative 0;	Mismatches 121;	Indels 6;	Gaps 1;

Qy	86	aggcattcctggaggaggaactccctcctgggttttactctaaagggtggcctggagcacg	145
Db	59	aggtagctctggaggggccagcacctcgggctctccgctgcaagggggcaaggactcca	118
Qy	146	gagaaaccattaatctcttaaggctcgaagaaggggggcgaagcagacacctgagctcca	205
Db	119	atgtgccctctccattcccgctcactcctgggggcaaaagggc-----gcaggccg	172
Qy	206	aactgcaggctggggatgaggtgtgcacatcaatgactgagctcggagcagctccagaa	265
Db	173	gagtgccgctggggcagctgggtcgtgagcatcgatggcgagaatcggggtgagctcacac	232
Qy	266	aggaggcagtttcctcctggtgaaaggatcctacaagacctcaggctgggtagtcgcgacgc	325
Db	233	acatcgagctcagaacaagatccgggctcggggagcgctcagcctgggcctcagca	292
Qy	326	tctcccacccggt	338

[illegible]

```

RESULT      8
US-09-124-238A-21
; Sequence 21, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
; TITLE OF INVENTION: Expression Systems
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS Word
; SEQ ID NO 21
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-21

```

Query Match	6.0%	Score 34.2;	DB 4;	Length 765;
Best Local Similarity	49.0%;	Pred. No. 0.12;		
Matches 124;	Conservative 0;	Mismatches 123;	Indels 6;	Gaps 1;
QY	86	aggcattctgaggaggagctccctggggttttactctaaagggtggcctggagcacg	145	
Db	74	aggtagtgctgaggggcccagcactctggggtctccggctgcgaagggggcaaggacttca	133	
QY	146	gagaaccattaatcattcttaagtgtagaagaagggggcaagcagacacctgagctcca	205	
Db	134	atgtgccctctccatttcccgctcacctctctggggcaagg-----ccgtgcaggcgcg	187	
QY	206	aactgcagcctggggatgaggtgtgcacataatgagtgactctgagcagctccagaa	265	
Db	188	gagtgcccgtaagtgactgctgagcatcgagggcgagaatgcgggtgacctcacac	247	
QY	266	aggagcgagtttccttgtagaaggatctcacaagacctcaggctggtagtggcgagcc	325	
Db	248	acatcgagctcagaaacaagatccggctgcggggagcgctcagcctgggctcaaca	307	
QY	326	tctcccaccgggt	338	
Db	308	gggccagccggt	320	

RESULT 9
 US-08-967-513-1
 : Sequence 1, Application US/08967513
 : Patent No. 5783436
 : GENERAL INFORMATION:
 : APPLICANT: Robert P. Hausinger
 : TITLE OF INVENTION: Mutant Urease and Method
 : TITLE OF INVENTION: of Use For Determination
 : TITLE OF INVENTION: of Urea
 : NUMBER OF SEQUENCES: 5
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Ian C. McLeod
 : STREET: 2190 Commons Parkway
 : CITY: Okemos
 : STATE: Michigan
 : COUNTRY: USA
 : ZIP: 48864
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 5.25 inch, 360 kb
 : MEDIUM TYPE: Storage
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: MS-DOS
 : SOFTWARE: Wordperfect 5.1
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/967,513
 : FILING DATE: 11-NOV-1997
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/687,645
 : FILING DATE: July 26, 1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Ian C. McLeod
 : REGISTRATION NUMBER: 20,931
 : REFERENCE/DOCKET NUMBER: MSU 4.1-309
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (517) 347-4100
 : TELEFAX: (517) 347-4103
 : TELEX: No. 57834366
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2400
 : TYPE: nucleotides
 : STRANDEDNESS: Single
 : TOPOLOGY: Linear
 : MOLECULE TYPE:
 : DESCRIPTION: cDNA
 : HYPOTHETICAL: No
 : ANTI-SENSE: No
 : ORIGINAL SOURCE:
 : ORGANISM: Klebsiella aerogenes
 : STRAIN: CG253
 : INDIVIDUAL ISOLATE:
 : CELL TYPE: N/A
 : FEATURE:
 : NAME/KEY: cDNA encoding mutant urease
 : NAME/KEY: h2190
 : LOCATION: Modification at position 1312 to
 : LOCATION: glutamine
 : IDENTIFICATION METHOD: Sequencing
 : OTHER INFORMATION:
 : US-08-967-513-1

	Query Match	5.9%	Score 33.8	DB 1	Length 2400
	Best Local Similarity	46.4%	Pred. NO. 0.3		
	Matches 110	Conservative 0	Mismatches 127	Indels 0	Gaps 0
QY	136	ctggagcagcgggaaccattaatcatctcaaggtcgaagaaggggcgaagcagacacc	195		
	127	CCCCCGTCGTCGGCCCTGATCAGCGCCTTTATTATGAAAGCGCCTCGGACGCGCAAAAGC	186		

IDENTIFICATION METHOD: Sequencing

US-08-166-316-1

Query Match	5.8%;	Score 33.4;	DB 1;	Length 1597;
Best Local Similarity	55.7%;	Pred. No. 0.32;		
Matches 64;	Conservative	0;	Mismatches 51;	Indels 0;
				Gaps 0;

Qy	86	agggaattcctggaggaggagctccctcctggggttttactataaagggtgcccctggagcag	145
Db	14	AAGTAGTCCTGGAGGGGCCAAGACCTTGGGGCTTCGGCTGAAGGGGCAAGGACTTCA	73
Qy	146	gagaacattaatcatctctaagtcgaagaagggggcaaacgacagacacccctgag	200
Db	74	ATGTGCCCTCTCAATTTCCCGCTCACCTCTGGGGGCAAGCGGCAGCGGAG	128

```

RESULT 12
US-08-948-705-5
; Sequence 5, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND
; TITLE OF INVENTION: TREATING COLON CANCER
; FILE REFERENCE: LOD-5506-JEL/NDH
; CURRENT APPLICATION NUMBER: US/08/948,705A
; CURRENT FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 2162
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-08-948-705-5

```

Query Match	5.8%	Score 33.4	DB 3	Length 2162
Best Local Similarity	50.7%	Pred. No. 0.38		
Matches 110	Conservative 0	Mismatches 101	Indels 6	Gaps 1
QY 115	ggttttaactaaagggtggcctgagacagcagagaaaccattaatctcttaagtgcaa	174		
Db 391	ggcctgagtgcgctggggcctggagtttgctggctctcatctccacctc	450		
QY 175	gaagggggaaagcagacacocctgagctccaaaactcaggctggggatgaggttctgac	234		
Db 451	aaagcggctcaggcagac-----agcgtgggtccaggtaggggacgagatggtccgg	504		
QY 235	atcaatgaggtgactctgagcagctccagaaagaggcaggtttccctgggtaaggatcc	294		
Db 505	atcaatggatattccatctctctctgtaaccatgaggaggtcatcaacctcattctgaacc	564		
QY 295	tacaagacctcaggctgggtagtcgcagcctctccc	331		
Db 565	agaagaaactgtgtccatcaagtgagacacatcgcc	601		

```

RESULT 13
US-08-948-705-4
; Sequence 4, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Chen, rao-Tseng
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND
; TITLE OF INVENTION: TREATING COLON CANCER
; FILE REFERENCE: LUD-5506-JEL/NDH
; CURRENT APPLICATION NUMBER: US/08/948,705A
; CURRENT FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 2236

```

```

; TYPE: DNA
; ORGANISM: Homo Sapiens
US-08-948-705-4

```

	Query Match	5.8s;	Score 33.4;	DB 3;	Length 2236;
	Best Local Similarity	50.7s;	Pred. No. 0.39;		
	Matches 110;	Conservative 0;	Mismatches 101;	Indels 6;	Gaps
QY	115	ggttttacttaaaggctggcctgagcagcagagaaccattaatcatctctaagggtcgaa	174		
Db	391	ggcttgagtgctggggcctggagtttggctgtgggtctctcatctcccaccctac	450		
QY	175	gaagggggcaagcagacacccctgagctccaaaactgcaggctggggatgaggttggc	234		
Db	451	aaagcgctcaggcagac-----agcgtcgggtccaggtaggggacgagatcgctcgg	504		
QY	235	atcaatgaggtgactctgagcagctccagaaagagcagtttccctggtgaagatcc	294		
Db	505	atcaatggataattccattctctctgtaccctagaggaggtcatcaacctcatctcgaaoc	564		
QY	295	tacaagaccctcaggctgggtagtcgcgagcctctccc	331		
Db	565	aagaaaactgtgtccatcaagaagtgagacatcgqcc	601		

```

RESULT 14
US-08-948-705-8
; Sequence 8, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND
; TITLE OF INVENTION: TREATING COLON CANCER
; FILE REFERENCE: LUD-5506-JEL/NDH
; CURRENT APPLICATION NUMBER: US/08/948,705A
; CURRENT FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 2289
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-08-948-705-8

```

[illegible]

```

; Sequence 6, Application US/08739485
; Patent No. 5863898
; GENERAL INFORMATION:
; APPLICANT: GOLI, Surya K.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/739,485
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0142 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1336 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
;
; US-08-739-485-6

```

```

Query Match          5.8%; Score 33.2; DB 2; Length 1336;
Best Local Similarity 54.0%; Pred. No. 0.34;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 65 aggaaggtacattatctgagagcattcctgagaggaagcctcctgaggttttactc 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 AGGAAGGCGGCGATCCCGACGCGTGATCCCGCGGCCCTCGGGGCTTCAGGC 128

QY 125 taaagggtgacctgagacaggaaccattatcatctcaaggtcgaagaagggggca 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 TCTCAGGGGCGCATGACTTCACCGAGCCTTGTGATCATCACCAGATTACACGGAAGCA 188

QY 185 aagcag 190
    ||| |||
Db 189 AGGCGG 194

```

Search completed: January 31, 2002, 19:50:03
Job time: 96 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:16:22 ; Search time 1599.5 seconds
(without alignments)
5909.899 Million cell updates/sec

Title: US-09-641-831-1
Perfect score: 573
Sequence: 1 atgatggagcactgaaga.....gagcctatttaagatga 573

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank1.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_om.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vl.*
30: em_hgo_hum.*
31: em_hgo_inv.*
32: em_hgo_rnd.*
33: em_hgo_hum.*
34: em_hgo_inv.*
35: em_hgo_rnd.*
36: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	573	100.0	573	6	AX088023 Sequence
2	323	56.4	327	6	AX088025 Sequence
3	323	56.4	402	6	AX088027 Sequence
4	253	44.2	156375	2	AP002859 Homo sapi
5	253	44.2	159488	2	AC025141 Homo sapi
6	209	36.5	6423	10	AF199421 Mus muscu
7	170	29.7	163335	2	AC046189 Homo sapi
8	97.4	17.0	7445	9	HSAPX1
9	85.8	15.0	6014	9	AB033028 Homo sapi
10	64.6	11.3	196542	9	AC002365 Homo sapi
11	45	7.9	156413	9	HSAL19E20 Homo sapi
12	44.8	7.8	39213	9	AC003047 Homo sapi
13	44.8	7.8	44400	9	AC090481 Homo sapi
14	43.6	7.6	173510	9	AL359272 Human DNA
15	37.8	6.6	2111	9	S39392 Homo sapien
16	37.8	6.6	3984	9	HUMCAP
17	37.4	6.5	1567	9	AF345905 Homo sapi
18	37.4	6.5	1669	9	AF345904 Homo sapi
19	37.4	6.5	1686	9	AF345906 Homo sapi
20	36.8	6.4	1485	10	AF002283 Mus muscu
21	36.8	6.4	329709	1	AP002997 Mesothelio
22	36.6	6.4	3220	9	BC008741 Homo sapi
23	36.6	6.4	3287	9	AF061258 Homo sapi
24	36.6	6.4	207079	2	AC015743 Homo sapi
25	36.6	6.4	209749	2	AC015564 Homo sapi
26	36.2	6.3	1696	10	AF095585 Rattus no
27	35.8	6.2	1724	9	AF265209 Homo sapi
28	35.8	6.2	1725	9	HUMEN15MA
29	35.8	6.2	1757	9	BC001093 Homo sapi
30	35.8	6.2	1896	9	RNU48247 Rattus norv
31	35.8	6.2	7218	6	166494 Sequence 14
32	35.8	6.2	134482	9	AC007335 Homo sapi
33	35.8	6.2	160460	2	AC020766 Homo sapi
34	35.6	6.2	112027	9	AC007006 Homo sapi
35	35.6	6.2	187727	2	AC064806 Homo sapi
36	35	6.1	2283	9	AB006955 Homo sapi
37	34.8	6.1	1282	5	SSA308531 Salmo sal
38	34.8	6.1	147197	2	AC091912 Homo sapi
39	34.8	6.1	171280	2	AC024186 Homo sapi
40	34.8	6.1	183349	2	AC027501 Homo sapi
41	34.8	6.1	237308	2	AC009593 Homo sapi
42	34.6	6.0	1191	10	RNRIT8
43	34.6	6.0	4542	3	AB036856 Homo sapi
44	34.4	6.0	1607	9	HSAL13766 Homo sapi
45	34.4	6.0	195630	2	AC087898 Mus muscu

ALIGNMENTS

RESULT 1

AX088023 573 bp DNA PAT 17-MAR-2001
LOCUS AX088023 Sequence 1 from Patent WO0114422.
DEFINITION AX088023
ACCESSION AX088023
VERSION AX088023.1 GI:13396951
KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (Bases 1 to 573)

AUTHORS

Turner C.A., Zambrowicz B., Friedrich G., Nehls M. and Sands A.T.

TITLE

Sequence derived from a human mammary gland cDNA library

JOURNAL

Patent: WO 0114422-A 1 01-MAR-2001;

FEATURES

source

1. 573
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT	91 a	76 c	96 g	64 t
ORIGIN	/dp_xiel=caxon:9606			

Query Match 56.4%; Score 323; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 7.4e-85;
Matches 323; Conservative 0; Mismatches 0; Indels

Qy	1	atgatgaggaccactgaagacttcacaagacctagtccacattaaactctaaacacggcc	60
Db	1	ATGATGAGGACCACTGAAGACTTCACAAGCCTAGTGCCACATTAAACTCTAAACACGGCC	60
Qy	61	accaagggaaggtacatttactgagagcctctcgagggagagctccctgggggtttt	120
Db	61	ACC AAGGGAAGGTACATTTTCTGTGAGGCATCTCTGAGGGAGGAGCTCCCTGGGGGTTTT	120
Qy	121	acttaaggggtggcctggagcagcggaaccattaatcatctctaaggtcgaagaagg	180
Db	121	ACTCTAAGGGTGGCCTGGAGCAGGAGACCAATTATCATCTCTAAGTTCGAAGAAGGG	180
Qy	181	ggcaagcagacacccctgagctcacaactcgaggtcgggatgaggttgtgcacataat	240
Db	181	GGCAAGCAGACACCCCTGAGCTCCAAACTGCAGCTGGGATGAGTGTGCACATCAAT	240

Oy	241	gagtgactctgcgacgtctcagaaggagcgatcttcccccggcgaaagagatccaccag	300
Dd	241	GAGTGACTCTGCACGAGGTCCAGAAAGGAGCGAGTTCTCCTGTGAAGAATCCTACAG	300
Oy	301	accctcagctcgtagtcgcag	323
Dd	301	ACCCTCAGCTCGTAGTCGCAG	323
RESULT 4			
AP002859		156375 bp	DNA
LOCUS			HNG 29-AUG-2000
DEFINITION			Homo sapiens chromosome 4 clone 225E15 map 4q16-q20, *** SEQUENCING
ACCESSION			AP002859
VERSION			AP002859.1 GI:9955382
KEYWORDS			HTG, HNGS, PHASE1.
SOURCE			Homo sapiens DNA, clone:225E15.
ORGANISM			Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE			1 (bases 1 to 156375)
AUTHORS			Tsal,S.F.
TITLE			Direct Submission
JOURNAL			Submitted (28-AUG-2000) Shih-Feng Tsal, National Yang-Ming University, Institute of Genetics, 15 Li-Kong St. Section 2, Pelou, Taipei, Taiwan 11221, Republic of China (E-mail:jympeisa@ym.edu.tw, URL:http://genome.ym.edu.tw/, Tel:886-2-28267043, Fax:886-2-28264930)
COMMENT			These sequences are draft human sequences, not finished sequences. These sequences are unordered pieces. And gaps between the contigs of the same clone are represented as 100 N. NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
			1 21185: contig of 21185 bp in length
			* 21186 21285: gap of 100 bp
			* 21286 22928: contig of 1643 bp in length
			* 22929 23028: gap of 100 bp
			* 23029 26285: contig of 3257 bp in length
			* 26286 26385: gap of 100 bp
			* 26386 30606: contig of 4221 bp in length
			* 30607 30706: gap of 100 bp
			* 30707 36924: contig of 6218 bp in length
			* 36925 37024: gap of 100 bp
			* 37025 40953: contig of 3929 bp in length
			* 40954 41053: gap of 100 bp
			* 41054 47843: contig of 6790 bp in length
			* 47844 47943: gap of 100 bp
			* 47944 57174: contig of 9231 bp in length
			* 57175 57274: gap of 100 bp
			* 57275 64994: contig of 7720 bp in length
			* 64995 65094: gap of 100 bp
			* 65095 76523: contig of 11429 bp in length
			* 76524 76623: gap of 100 bp
			* 76624 88306: contig of 11683 bp in length
			* 88307 88406: gap of 100 bp
			* 88407 101162: contig of 12756 bp in length
			* 101163 101262: gap of 100 bp
			* 101263 115436: contig of 14174 bp in length
			* 115437 115536: gap of 100 bp
			* 115537 133350: contig of 17814 bp in length
			* 133351 133450: gap of 100 bp
			* 133451 156375: contig of 22925 bp in length.
FEATURES			Location/Qualifiers
SOURCE			1..156375
			/organism="Homo sapiens"
			/db_xref="taxon:9606"

Query Match	44.2%	Score 253;	DB 2;	Length 156375;
Best Local Similarity	100.0%	Pred. No. 5,3e-64;		
Matches 253; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
BASE COUNT	45683 a	32392 c	31715 g	45174 t 1411 others
ORIGIN				
Query Match	/chromosome="4"			
	/map="4q16-q20"			
	/clone="225E15"			
321 cagcctccccccacgcytaccgtttagcttagctcgtagtttgaccctcaacatccccagaagat	380			
Db 31166 CAGCCTCTCCCAACCGGTACGTGTAGCTCGAGTTTGACCCCAACATCCCAAGAGAT	31225			
381 gctccttaggactcgaaacctttagtctcttactctgtatgtagcgcagatgagtgag	440			
Db 31226 GCCTCTAGACACAGCAACCTCATTTAGTGTCTCTACTGCTGAGAGAGCCATGAGTGAG	31285			
441 cgtgcaccacacttggtgtagaagtggtgtctcaagtcaccgttcacatggagcagcatgagcc	500			
Db 31286 CTGTCCACACACTTGGGTGAGTGTGTCTCCACCTCCCACTGGGACAGATAGGCC	31345			
501 acagaaaggttgatctatcaccacccacagacacacatgacgaattccaagaagccta	560			
Db 31346 ACAGAAAGGTGTATCTACCCCAACCCAGACACAAACATGCAGAAATTTCAAAGAGCCTA	31405			
561 tttaagtagatga	573			
Db 31406 TTTAAGTAGATGA	31418			
RESULT 5				
AC025141	AC025141	159488 bp	DNA	HTG 23-SEP-2000
LOCUS	Homo sapiens clone RP11-224D4,	WORKING DRAFT	SEQUENCE, 8 unordered	
DEFINITION	pieces			
AC025141	AC025141.4	GI:10280891		
VERSION	AC025141.4	GI:10280891		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 159488)			
JOURNAL	Barren,B., Linton,L., Nusbaum,C. and Lander,E.			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 159488)			
	Barren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,			
	Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,			
	Boguski,M., Bouckgalter,B., Brown,A., Burkett,G.,			
	Campbell,A., Castle,A., Choepey,Y., Colangelo,M., Collins,S.,			
	Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,			
	Dodgson,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gege,D.,			
	Gallagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,			
	Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,			
	Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,			
	Klein,J., Labouque,K., Lamazares,R., Landers,T., Lehoczy,J.,			
	Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,			
	McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R.,			
	Meldrum,J., Meneus,L., Minova,T., Miranda,C., Mleaga,V., Morrow,J.,			
	Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,			
	O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,			
	Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,			
	Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,			
	Stange-Thomann,N., Stojanovic,N., Sudranthan,A., Talamas,J.,			
	Tejafes,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,			
	Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.,J.,			
	Young,G., Zainoun,J., Zimmer,A. and Zody,M.			
TITLE	Direct Submission			
JOURNAL	Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome			
	Research, 320 Charles Street, Cambridge, MA 02141, USA			

COMMENT

On Sep 23, 2000 this sequence version replaced gi:8077070.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7938
Center clone name: 224_D4

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 155108 bases at least Q40
Consensus quality: 157181 bases at least Q30
Consensus quality: 158099 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 158788; sum-of-contigs
Quality coverage: 6.8 in Q20 bases; agarose-fp
Quality coverage: 6.4 in Q20 bases; sum-of-contigs

* NOTE: this is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 11564: contig of 11564 bp in length
* 11565 11664: gap of 100 bp
* 11665 14943: contig of 3279 bp in length
* 14944 15043: gap of 100 bp
* 15044 15938: contig of 4395 bp in length
* 15939 19538: gap of 100 bp
* 19539 31843: contig of 12305 bp in length
* 31844 31943: gap of 100 bp
* 31944 98274: contig of 66331 bp in length
* 98275 98374: gap of 100 bp
* 98375 114530: contig of 16176 bp in length
* 114531 114650: gap of 100 bp
* 114651 147429: contig of 32779 bp in length
* 147430 147529: gap of 100 bp
* 147530 159488: contig of 11959 bp in length.

FEATURES
Location/Qualifiers

1..159488
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-224D4"
/clone_lib="RP11-224D4"
1..11564
/note="assembly_fragment
clone_end:SP6
vector_side:left"
11665..114943
/note="assembly_fragment"
15044..119438
/note="assembly_fragment"
19539..31843
/note="assembly_fragment"
31944..98274
/note="assembly_fragment"
98375..114530
/note="assembly_fragment"
114651..147429
/note="assembly_fragment"
147530..159488
/note="assembly_fragment
clone_end:T7
vector_side:right"

BASE COUNT

46031 a 33033 c 33024 g 46700 t 700 others
ORIGIN

Query Match 44.2% Score 253; DB 2; Length 159488;
Best Local Similarity 100.0%; Pred. No. 5.3e-64;
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 321 cagcctctccccaccggtcactgttagctgctgagtttgacctcaacatccccagagat 380
|||||
Db 2903 CAGCCTCTCCCCACCGGTCACTGTTAGCTCGAGTTTGACCTCAACATCCCAGAGAT 2962
Qy 381 gcctctaggactggaacctcatttagtctctactgctgtagcgcctagtgagag 440
|||||
Db 2963 GCCCTCTAGGACTCGAACCTCATTTAGTGTCTCTACTGTGATGGACGCATGAGTGGAG 3022
Qy 441 ctgtcgaccacctgggtggaagtgtgtctccagctccaccctggcgagacagatggcc 500
|||||
Db 3023 CTGTGACCACCTTGGGTGAAGTGTGTCTCCAGTCCCACTGGGACGACGATGGCC 3082
Qy 501 acagaaaggttcttctaccacccaccagcacacacatgcagaaatttcaaaagagccta 560
|||||
Db 3083 ACAGAAAGGTTGTATCTTACCCACCACACACATGCAGAAATTTCAAAAGAGCCTA 3142
Qy 561 tttaagtagatga 573
|||||
Db 3143 TTTAAGTAGATGA 3155

RESULT 6

AF199421 6423 bp mRNA ROD 13-DEC-1999
LOCUS Mus.musculus PDZ domain actin binding protein Shroom mRNA, complete
DEFINITION cds.
ACCESSION AF199421
VERSION AF199421.1 GI:6467989
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 6423)
AUTHORS Hildebrand,J.D. and Soriano,P.
TITLE Shroom, a PDZ domain-containing actin-binding protein, is required
for neural tube morphogenesis in mice
JOURNAL Cell 99 (5), 485-497 (1999)
MEDLINE 20055594
PUBMED 10589677
REFERENCE 2 (bases 1 to 6423)
AUTHORS Hildebrand,J.D. and Soriano,P.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Basic Sciences, Fred Hutchinson Cancer
Research Center, 1100 Fairview Avenue North, Seattle, WA
98109-1024, USA

FEATURES

Location/Qualifiers
1..6423
/organism="Mus musculus"
/db_xref="taxon:10090"
233..6193
/note="shrm"
/codon_start=1
/product="PDZ domain actin binding protein Shroom"
/protein_id="AAF13269.1"
/db_xref="GI:6467990"
/translation="MKTPENLEEPSATNPSTPTFRFVYLEALLEGAPWGFTLKGG
LERGEPLIISKIEEGKADSVSGLOAGDEVTHINEVALSPRRRAVSLVKGSKYKTLR
LVVRDVCAAPGHADPGTSGKSLSELLTCSPOHRKATWSGVKRLKORCSPPAPRPH
SWHTTKFGETQPDVSMOISOGCTGPPPHQSHSSSSSDLSNHYALRRSPDCSS
QGSMELEPSSGYPCHLLSPAKTSSIDQLGHLNKRDSAYSSSTSSSIFPPPG
GSARSGMDVISARGLLGEMQADINRYKTYDTRGVSSVEYVNPASALLQGRD
AHASADSGCAKHSIPKGGTTPSPSQOCSGSLTADNLNPKAGAPLPPTRSDSY
AAFRHRERPSWSSLDHFRFCRPQTNSSGSKPTFAEDQLHTVPPRSPSPSPVSKSH
NYTKAQPGQPLLPTGIYVPSPPEPFAQVPQSPVSSNGTVYFALVKESGYTAAGTC

TITLE	Direct Submission
JOURNAL	Submitted (20-AUG-1997) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	4 (bases 1 to 196542)
AUTHORS	Chiu, M.W.
TITLE	Direct Submission
JOURNAL	Submitted (04-SEP-1997) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA
REFERENCE	5 (bases 1 to 196542)
AUTHORS	Worley, K.C.

```

/Note="Similar to sequence with GenBank Accession number
236840"
repeat_region      15384..15565
                    /rpt_family="MER58B"
repeat_region      15656..15956
                    /rpt_family="AluSg"
repeat_region      16091..16691
                    /rpt_family="L1MB7"
                    /join(<17023..11776,34350..34481,37731..40071,41563..41663,
75548..76243,80507..81058,82568..82739,88014..88286,
90044..92814)
                    /gene="HSAAPL"
                    /product="APPL"
                    /note="human homolog of Xenopus laevis APX gene, similar
to sequence encoded by GenBank Accession Number X83543"
                    /codon_start=3
                    /product="APXL"
                    /protein_id="AAC32592.1"
                    /db_xref="GI:1773381"
                    /translation="IEGSKAAAYDKILLAGEIYVINDIGLSEGRQEAICLVKSGHKT
LKVYKRSEGLWPHSHHATKESDSHPELAASPTSTSGCPMSGRHSASSHDLS
SSMOQNTQRLDHRSSLSGVSDLDHPSRLSVAKSNSIDHLSHSHKRDASVSEFT
SSSTPDHLSKADTSASENIIYVGLWEAPYQGGQAQADDPQSEKELCPFRVRS
GDSGPRPEYNAEKLAPGRSNFGVWYVDKKAPSSPPPPPLRSQFAATKS
HEKAGDPYSESAQAHOFTALAOAPRGDRPELDWRASHPGLSGGPGCPQ
EADADGSPKSKDAASRLQASISDVRFQSPHSGHPLYSHPSLCADLQOE
GAASFQNDSPQVRLSSCDOKLGGDPPRCVQAOALMGACMPSPDALALE
SLPPTVQGSPPRHLPQEPGPPDAETRCPLDKGAGCCASGAPPRASAEKASQ
RLAASITWADGSSRICPEETPLHSLTQEGRRPSSPESATPPPPPDVAGCTR
RSDRATILRNELIOMHRAKIQRSSTVALTAGEMDGTGRRAALGGTQDPLAGT
YKDHLEQAQVRLATSRFRKRDIDNPDGLYSELEHMGDPDLYVPHWEALQAP
STSGGPHPRJGGRRTAEOKIKSGPEKANEGLRGYSPHQHPTSEDTVGTFLN
DMAKFEETSKYVPPRPAKQKILSTIPDKPRPTAGCTGPEWSTTSLGSLN
POHVGRKSSPSDHYKQEAVELRKQGDGEPREELPSVRAEBCSTROADQ
CREGSPSQHPQKAPNPPTSELSHCRGAPELPRREGSAGLTPROTKRISSESTP
ADIGRAOSPGSPILHARGDSWVSALSKRPAPQRPDPKATCERGSQHVSDA
PQVIGRPPPTSPASIDYVARLSHSBVSFSSAOPQDTPKATCERGSQHVSDA
SRPLPALLPQQHRLQOTAMETSRSPQAPKLDKPLLIQDSDSTRIEFVM
DNNTYKMPPIKIVHESQPEKESROSLACPAEPALPHGLKDKDITKSTSEQFVS
FLITYROGAPEPAPRAQAPQPLGTQVPEKDRCTSPGICSYMAKAKTYDCKSE
ELAREIVDKSLADILPSVKITTMDEGIFPEDHELDQOQRKRLKPTPBR
SREERKEEVSVAVALATNSYVSTSAKAEELLIMKDLQOEHEDESGSLDLDL
SVYKQRLISIKQLVLEARESELIEDVOANTVGAEEAIVKGVCKSEFEKFEF
IGDLDVYVILLISGRILARVNALNNLDGASPDROSILEKORVILIOHEAKELK
ENDRRERIVFDILANTLSESLADYEHYKMSALITIORELEDIHIGEEQLKCLL
DSIQPERGR"
repeat_region      complement(17976..18074)
                    /rpt_family="Hy3"
repeat_region      18162..18473
                    /rpt_family="AluJb"
repeat_region      19636..19717
                    /rpt_family="MER21A"
repeat_region      complement(19798..19937)
                    /rpt_family="L1MB3"
repeat_region      complement(20181..20327)
                    /rpt_family="FRAM"
repeat_region      complement(22367..22401)
                    /rpt_family="AT_rich"
repeat_region      22406..22461
                    /rpt_family="GGAAn"
                    complement(22462..22627)
                    /rpt_family="AluJo"

```

```

repeat_region      complement(22638..22933)
                    /rpt_family="AluSx"
repeat_region      complement(26881..26957)
                    /rpt_family="MER5A"
repeat_region      complement(27276..27762)
                    /rpt_family="POLY_A"
repeat_region      complement(27876..27977)
                    /rpt_family="AluJo/FRAM"
repeat_region      complement(28048..28069)
                    /rpt_family="AT_rich"

Query Match      11.38; Score 64.6; DB 9; Length 196542;
Best Local Similarity 65.7%; Pred. No. 3.9e-08;
Matches 94; Conservatve 0; Mismatches 49; Indels 0; Gaps 0;

QY 175 gaaggggcaagaagacacccctagctccaaatgcagctgggagatgagttgac 234
    |||||
Db 17028 GAAGAGGCGCTAAGACCGCGCGCTGACAGATTACTGGGATGAGATGATCGTGGC 17087

QY 235 atcaatgagtgactctgagcagctccagaagaaggagcagttccctgggaaagatcc 294
    |||||
Db 17088 ATCAATGACATTGCTCTCTCAGGCTTTAGACAGAGGAGATTGCTGTGTGAAGGGGTCC 17147

QY 295 tacaagaccctcagctgctgagt 317
    |||||
Db 17148 CATAGACCTCGAAGCTGCTGCT 17170

```

```

RESULT 11
LOCUS      HSA119E20
DEFINITION Human DNA sequence from clone RP11-119E20 on chromosome
            Xp11.21-11.23 Contains part of the gene for KIAA1202 protein, ESTs,
            STSs and GSSs, complete sequence.
ACCESSION  AL121865
VERSION    AL121865.7 GI:8979790
KEYWORDS   HTG; KIAA1202.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 156413)
            Wilson,S.
            Direct Submission
            Submitted (27-Jul-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Jul 8, 2000 this sequence version replaced gi:8977862.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-119E20 is from the library RP11-11.1 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACE3.6

```

```

FEATURES             This sequence is the entire insert of clone RP11-119E20.
Source               1. .156413
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="X"
                    /map="p11.21-11.23"
                    /clone="RP11-119E20"
                    /clone_lib="RPC1-11.1"
                    8. .593
misc_feature
repeat_region       /note="match: GSS: Em:AQ350374"
                    1965. .2081
                    /note="L2 repeat: matches 2602. .2708 of consensus"
                    2374. .2464
                    /note="L2 repeat: matches 2610. .2686 of consensus"
                    2607. .2755
                    /note="MIR repeat: matches 27. .176 of consensus"
                    join(<3161. .3312,60790. .60924,63430. .65920,71423. .71484,
                    90914. .91717,96285. .96465,100563. .100832,102134. .103400)
                    /genes="ball19E20.1"
                    /note="match: cDNAs: Em:AB033028 Em:X83543 Em:AB040914"
                    /evidence=not_experimental
                    /product="ball19E20.1 (KIAA1202 protein)"
                    3161. .103400
                    /gene="ball19E20.1"
                    join(<3161. .3312,60790. .60924,63430. .65920,71423. .71484,
                    90914. .91717,96285. .96465,100563. .100832,102134. .102403)
                    /gene="ball19E20.1"
                    /note="match: proteins: Tr:Q9ULL8 Tr:Q9QXW9"
                    /codon_start=1
                    /evidence=not_experimental
                    /product="ball19E20.1 (KIAA1202 protein)"
                    /protein_id="CAC08472.1"
                    /db_xref="GI:12314062"
                    /translation="IEDGKAALSQKRTGDELVNINGTPLYGSRQALILIKGSPRI
                    LKLVRRNPVSPRHSHVAKLLEGCPAAATMHPPESEAFSLSWHSGCNTSDVCVOM
                    CPLEHSTKESIGSMESLEQPGQATYVSHLLPIDONMIFNQRDSAYSSFSASSNA
                    SDALSLRPEASTDCIMOGPGPTKAPSRPNVAETSGGSRRTNGHLTPSSQMSRP
                    PQEYQSGAKAVRGPQPVVRDLSQAASRQLLNGEQRRASEFVPLPQKEKLSLEP
                    VLPARNRRCCLSGDVPVYSEGCNCFESQPPSSQSGSEHLLMOASTKAVGSPKAC
                    DRASSVNPPLNEASAEAKAASFGPRPHLIGTGHRSAPQELLASHLQHVHLDTRG
                    KGVELDPVQGHQWTLSPHSHGKKSPPCPTGGTHDQSSGSKERTKROVDVRSILGH
                    SQSSPPHQBAGHPEKSGKGLDNRTSRAASELANQOAPSSAGSLVOQATDCSTTKAA
                    SKPLQGLCTKEPVEETQPPSPPLTASNTSLSSCKPPSPROKLNKSMMLRAG
                    SSECLQAPSHESRGLGRLISPGORPQSGSLGNTWWKAPDPPSSDPEKAHAGCV
                    RGGHWRMSPHNSQPLVAAAMEGSPNPKELKASTAQAGEDAILLPADRRKKFEE
                    SSKSLSTSLPGLTTHSNKFTQRPAPIDONTQPMSSSCLRELRRHPMQSYHSADQPY
                    HATDQSYHNSPLQSETPYSECFASGLENSMCKPLHCGDFDVRTCSYSCSVQGA
                    LVHDPICYSGEICPALLKRNMPNVCYNCRHHOCIRCSVCYHNPQHSALDLSLAP
                    GNTWPKRKTUVEFFQDKWNPITGNKRTSQSGREMAHAKTSFWMATPHPCLENSLAP
                    LSSYRAISLIDLIGDFKHALKSEFTSVYEEGSSLASMPHLRSRAFSESHLSLAPQ
                    TRAWGRRRLFSKGDQTSDDLGAARKKAPPPRPPPNWERYLRFRAAQOQOQOQ
                    OKQOEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFE
                    LFGHLEGRSQGSQVPAQOESFALHSSDFLPIIRHLSQSQPEQAQPPCYTIGLWR
                    TSGOATESAKOEHFQFPSPGAPITSYSAYNISVAKELNKLKQDPEMAEIGL
                    GREEVDHELAQKIQILSISRLKSVLRQAORGLIEDINANSALLKEEVEANLKVCKS
                    NFEFKHFLVDRLKRVNLLSLGRLARVALNALSIDSEANOEKLVLEKKQQTGQ
                    LADAKELHVDRLKRVNLLSLGRLARVALNALSIDSEANOEKLVLEKKQQTGQ
                    EPQKLKRESLILGPSN"
                    4135. .4491
repeat_region       /note="L2 repeat: matches 1077. .1444 of consensus"
                    4494. .5509
                    /note="L1M4 repeat: matches 5256. .6300 of consensus"
                    5520. .5813
                    /note="AluX repeat: matches 1. .295 of consensus"
                    5824. .6234
                    /note="L2 repeat: matches 1460. .1940 of consensus"
                    7604. .7711
                    /note="L2 repeat: matches 2021. .2124 of consensus"
                    8125. .8413
                    /note="AluX repeat: matches 1. .289 of consensus"
                    8645. .9271
                    /note="L1M4 repeat: matches 121. .1333 of consensus"
                    9295. .9446
                    /note="L1 repeat: matches 3150. .3772 of consensus"
                    9878. .10064
                    /note="L1M4A repeat: matches 6100. .6293 of consensus"
                    10305. .10396
                    /note="L2 repeat: matches 2585. .2686 of consensus"
                    10404. .10478
                    /note="MER58A repeat: matches 127. .209 of consensus"
                    10479. .10831
                    /note="L1M4A repeat: matches 5669. .6038 of consensus"
                    10864. .11074
                    /note="MER58B repeat: matches 88. .326 of consensus"
                    11083. .11174
                    /note="L2 repeat: matches 2585. .2686 of consensus"
                    11683. .11813
                    /note="MER58A repeat: matches 1. .118 of consensus"
                    11809. .11936
                    /note="MER58A repeat: matches 97. .224 of consensus"
                    12188. .12303
                    /note="L2 repeat: matches 2594. .2709 of consensus"
                    12293. .12460
                    /note="MIR repeat: matches 34. .200 of consensus"
                    12512. .12626
                    /note="L2 repeat: matches 2628. .2744 of consensus"
                    12764. .13076
                    /note="AluY repeat: matches 1. .305 of consensus"
                    14579. .15021
                    /note="MLTIC repeat: matches 6. .466 of consensus"
                    15027. .15862
                    /note="L2 repeat: matches 1396. .2381 of consensus"
                    15863. .16314
                    /note="MLT2B repeat: matches 1. .444 of consensus"
                    16315. .16691
                    /note="L2 repeat: matches 2381. .2710 of consensus"
                    16933. .16960
                    /note="MIR repeat: matches 194. .261 of consensus"
                    16975. .17215
                    /note="L2 repeat: matches 1566. .1810 of consensus"
                    17250. .18409
                    /note="L2 repeat: matches 1612. .2750 of consensus"
                    18410. .18725
                    /note="L1M4C repeat: matches 7678. .7977 of consensus"
                    18465. .19300
                    /note="L2 repeat: matches 1412. .1615 of consensus"
                    19301. .19536
                    /note="L1M4 repeat: matches 6056. .6294 of consensus"
                    19537. .20202
                    /note="L2 repeat: matches 726. .1412 of consensus"
                    21281. .21390
                    /note="L2 repeat: matches 2636. .2750 of consensus"
                    23408. .23750
                    /note="L2 repeat: matches 1654. .2001 of consensus"
                    23752. .24002
                    /note="L1P13 repeat: matches 5929. .6152 of consensus"
                    24016. .24136
                    /note="L2 repeat: matches 2332. .2458 of consensus"
                    24333. .25345
                    /note="L1M4 repeat: matches 4144. .5147 of consensus"
                    25339. .26679
                    /note="L1M4C repeat: matches 1320. .2354 of consensus"
                    26687. .28148
                    /note="TIGGER1 repeat: matches 1. .1465 of consensus"
                    28149. .28979
                    /note="L1P7 repeat: matches 5315. .6145 of consensus"
                    28997. .29497
                    /note="TIGGER1 repeat: matches 1465. .1926 of consensus"
                    29498. .29764
                    /note="AluJb repeat: matches 1. .271 of consensus"
                    29765. .30244
                    /note="TIGGER1 repeat: matches 1926. .2418 of consensus"
                    30258. .31366
                    /note="L1M4C repeat: matches 121. .1333 of consensus"

```



```

misc_feature      11146..11196
                    /note="low coverage"
repeat_region     11203..11335
                    /rpt_family="FRAM"
repeat_region     11351..11652
                    /rpt_family="AluSx"
repeat_region     12228..12254
                    /rpt_family="AT-rich"
repeat_region     12257..12556
                    /rpt_family="AluSx"
repeat_region     12557..12590
                    /rpt_family="AT-rich"
repeat_region     12704..12994
                    /rpt_family="AluJo"
repeat_region     12995..13020
                    /rpt_family="GA)n"
repeat_region     13023..13093
                    /rpt_family="MER58A"
repeat_region     13667..13854
                    /rpt_family="AluSx"
repeat_region     14127..14233
                    /rpt_family="MER5A"
repeat_region     14368..14670
                    /rpt_family="AluSx"
repeat_region     15204..15494
                    /rpt_family="AluJb"
repeat_region     16134..16246
                    /rpt_family="MER4_internal"
repeat_region     16250..16538
                    /rpt_family="AluJb"
repeat_region     16660..16775
                    /rpt_family="AluJb"
repeat_region     17520..17805
                    /rpt_family="AluJb"
repeat_region     18199..18272
                    /rpt_family="MIR"
repeat_region     19058..19350
                    /rpt_family="AluJo"
repeat_region     19372..19584
                    /rpt_family="AluJb"
repeat_region     19599..19778
                    /rpt_family="MER58A"
repeat_region     19726..19778
                    /note="low coverage"
repeat_region     19900..20084
                    /rpt_family="AluJo"
repeat_region     20211..20364
                    /rpt_family="MER58A"
repeat_region     20280..20314
                    /rpt_family="POLY_A"
repeat_region     20415..20468
                    /rpt_family="FAM"
repeat_region     20472..20771
                    /rpt_family="AluY"
repeat_region     21254..21353
                    /rpt_family="AluJb"
repeat_region     21386..21447
                    /rpt_family="MTID"
repeat_region     22090..22391
                    /rpt_family="AluSx"
repeat_region     22393..22601
                    /rpt_family="L2"
repeat_region     23070..23235
                    /rpt_family="AluJb"
repeat_region     23484..23618
                    /rpt_family="AluJo"
repeat_region     23620..23920
                    /rpt_family="AluY"
repeat_region     23921..24091
                    /rpt_family="AluJo"
repeat_region     24131..24432
                    /rpt_family="AluJb"
repeat_region     24551..24787
                    /complement(24551..24787)

```

```

repeat_region     /rpt_family="AluJo"
                    25233..25550
repeat_region     /rpt_family="AluSp"
                    25566..25866
repeat_region     /rpt_family="AluJb"
                    25867..25901
repeat_region     /rpt_family="(GAAA)n"
                    /complement(26936..27339)
repeat_region     /rpt_family="LPR12"
                    /complement(27350..27573)
repeat_region     /rpt_family="LMD2"
                    28409..28490

```

```

Query Match      7.8%; Score 44.8; DB 9; Length 39213;
Best Local Similarity 69.3%; Pred. No. 0.028;
Matches 61; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 83 tggaggcattcttggaggagagctccctgggttttacttaagggtgacctgagc 142
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29327 TGGAGTGCAGCTGAGCGGCGGCCCTGGGCTTACCTGAGGCGCGCGAGC 29386

QY 143 acggagaaccattaatcatctcttaaggt 170
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29387 ACGGCGCGCGCTGCTCATCACAAGGT 29414

```

RESULT 13

```

AC090481 44400 bp DNA PRI 01-APR-2001
LOCUS Homo sapiens Xp BAC RP11-8H7 (Roswell Park Cancer Institute Human
AC090481 complete sequence.
AC090481
AC090481.9 GI:13491204
VERSION HTG.
KEYWORDS human.
SOURCE human.

```

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Brundage,E., Buhay,C.,
Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z.,
Cox,C., David,R., Delgado,O., Deshazo,D., Di,W., Ding,Y.,
Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C.,
Fragueto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Garcia,C.,
Gorell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Jackson,L.E.,
Hodgson,A., Hogue,M., Holloway,C., Hosak,H., Kong,Y., Kovar,C.,
Jackson,L., Jones,M., Kelly,S., Kondejewski,N., Logan,O.,
Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Martin,R., Martinez,C.,
Lozano,R.J., Lu,J., Lucier,R., Martin,R., McLeod,M.P.,
Mei,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R.,
Nguyen,N., Nguyen,S., Osval,G., Rampell,L.R., Parish,B., Paxton,S.,
Payton,B., Perez,L., Pu,L.L., Quiles,M., Rashid,N.D., Reiter,D.,
Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H.,
Simon,N., Sparks,A., Stamps,A., Sugang,R., Tabor,P., Taylor,T.,
Vasquez,L., Vinson,R., Vo,Q., Wabwah,M., Watlington,S., Taylor,T.,
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

```

REFERENCE

```

AUTHORS
2 (bases 1 to 44400)
Unpublished
2 (bases 1 to 44400)
Worley,K.C.
Direct Submission
Submitted (24-FEB-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 44400)
Worley,K.C.
Direct Submission
Submitted (31-MAR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 44400)

```



```

repeat_region      /rpt_family="L1MC5"
repeat_region      complement(9676..9885)
repeat_region      /rpt_family="MER58B"
repeat_region      9886..10166
repeat_region      /rpt_family="AluJb"
repeat_region      complement(10167..10258)
STS               /rpt_family="MER58B"
STS               10294..10455
/standard_name="OAL-CA"

Query Match
Best Local Similarity 69.3%; Pred.No.0.028; Length 44400;
Matches 61; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 83 tggaggcattctggggggaggagctccctgggttttactctaaaggtggcctggagc 142
Db 32491 TGGAGGTGACGCTGAGCGGGCGGCCGCTTACCTTGAGGGCGGCCGCGAGC 32550
QY 143 acggagaaccattaatcatctctaagt 170
Db 32551 ACGGCGAGCGCTGCTCATCACCAGGT 32578

RESULT 14
AL359272/c
LOCUS              AL359272      173510 bp      DNA      PRI      12-OCT-2000
DEFINITION         Human DNA sequence from clone RP11-554P16 on chromosome X. Contains
                    the first coding exon of the gene KIAA1202, STSS, GSSs and a CpG
                    island, complete sequence.
ACCESSION          AL359272
VERSION            AL359272.9  GI:9650579
KEYWORDS            HTG; CpG island; KIAA1202.
SOURCE             human.
ORGANISM            Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Bird,C.
Direct Submission
Submitted (28-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 1, 2000 this sequence version replaced gi:9588592.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Swi., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP11-554P16 is from the library RP11-11.2 constructed at the
Roswell Park Cancer Institute by the group of Pieter de Jong. For
further details see http://bacpac.med.buffalo.edu/
VECTOR: pBACs3.6
This sequence is the entire insert of clone RP11-554P16.
Location/Qualifiers
1..173510
/organism="Homo sapiens"
/db_xref="taxon:9606"

repeat_region      /chromosome="X"
repeat_region      /clone="RP11-554P16"
repeat_region      /clone_lib="RPC1-11.2"
repeat_region      2..1192
repeat_region      /note="L1MC repeat: matches 1913..2098 of consensus"
repeat_region      7..674
repeat_region      /note="match: GSS: Em:AQ373815"
repeat_region      195..511
repeat_region      /note="AluSp repeat: matches 1..312 of consensus"
repeat_region      543..845
repeat_region      /note="AluSx repeat: matches 1..304 of consensus"
repeat_region      950..1127
repeat_region      /note="L1MC repeat: matches 1968..2136 of consensus"
repeat_region      1190..2306
repeat_region      /note="L1M4b repeat: matches -180..888 of consensus"
repeat_region      2425..3343
repeat_region      /note="L1M4 repeat: matches 3078..4092 of consensus"
repeat_region      3344..3717
repeat_region      /note="MSTD repeat: matches 1..394 of consensus"
repeat_region      3718..4412
repeat_region      /note="L1M4 repeat: matches 4092..4274 of consensus"
repeat_region      4413..4452
repeat_region      /note="THE1B repeat: matches 1..42 of consensus"
repeat_region      4453..4763
repeat_region      /note="AluSx repeat: matches 1..310 of consensus"
repeat_region      4764..5111
repeat_region      /note="THE1B repeat: matches 42..364 of consensus"
repeat_region      5112..5774
repeat_region      /note="L1M4 repeat: matches 4274..5344 of consensus"
repeat_region      5775..6080
repeat_region      /note="AluSg repeat: matches 1..307 of consensus"
repeat_region      6081..6296
repeat_region      /note="L1M4 repeat: matches 5344..5581 of consensus"
repeat_region      6299..6923
repeat_region      /note="L1PA10 repeat: matches 5525..6165 of consensus"
repeat_region      7394..7652
repeat_region      /note="L1M4 repeat: matches 2120..2383 of consensus"
repeat_region      7652..10378
repeat_region      /note="L1PA4 repeat: matches 3414..6142 of consensus"
repeat_region      10369..10878
repeat_region      /note="L1M4 repeat: matches 2350..2879 of consensus"
repeat_region      10875..11250
repeat_region      /note="L1 repeat: matches 4741..5131 of consensus"
repeat_region      11247..11619
repeat_region      /note="L1M7 repeat: matches 5903..6284 of consensus"
repeat_region      11642..11764
repeat_region      /note="L1M4 repeat: matches 5351..5475 of consensus"
repeat_region      11764..11943
repeat_region      /note="L1 repeat: matches 3846..4036 of consensus"
repeat_region      12269..13359
repeat_region      /note="L1M3 repeat: matches 6659..7739 of consensus"
misc_feature       complement(13126..13490)
repeat_region      /note="match: STS: Em:HS3082G9"
repeat_region      13430..13497
repeat_region      /note="L1M4 repeat: matches 1995..2062 of consensus"
repeat_region      13498..13803
repeat_region      /note="AluJb repeat: matches 3..307 of consensus"
repeat_region      13804..15085
repeat_region      /note="L1M4 repeat: matches 720..1995 of consensus"
repeat_region      15191..15376
repeat_region      /note="L1M4 repeat: matches 408..602 of consensus"
repeat_region      15565..16035
repeat_region      /note="L1M4 repeat: matches -252..225 of consensus"
repeat_region      16058..16114
repeat_region      /note="L2 repeat: matches 2686..2745 of consensus"
repeat_region      16128..16369
repeat_region      /note="L1M10 repeat: matches 5928..6318 of consensus"
repeat_region      17285..17533
repeat_region      /note="match: GSS: Em:AQ142071"
repeat_region      17455..17899
repeat_region      /note="match: GSS: Em:AQ134667"
repeat_region      17875..18018
repeat_region      /note="MIR repeat: matches 96..252 of consensus"

```

```

repeat_region 18979..19286
/note="AluJb repeat: matches 2..308 of consensus"
repeat_region 19391..19823
/note="L2 repeat: matches 2284..2708 of consensus"
repeat_region 20320..20610
/note="AluJx repeat: matches 1..289 of consensus"
misc_feature 21916..22121
/note="match: STS: Em:G03829"
repeat_region 23553..23862
/note="MER3 repeat: matches 1..307 of consensus"
repeat_region 23946..24118
/note="MER3 repeat: matches 77..359 of consensus"
repeat_region 24356..24475
/note="L2 repeat: matches 2624..2750 of consensus"
repeat_region 24929..25125
/note="MIR repeat: matches 23..256 of consensus"
repeat_region 25251..25386
/note="MER63 repeat: matches 2..768 of consensus"
repeat_region 25604..25656
/note="L2 repeat: matches 2657..2709 of consensus"
repeat_region 26047..27589
/note="LIMB8 repeat: matches 4591..6159 of consensus"
repeat_region 27530..27886
/note="AluJx repeat: matches 1..297 of consensus"
repeat_region 27887..28214
/note="LIMB8 repeat: matches 4271..4591 of consensus"
repeat_region 28258..29096
/note="LIMB8 repeat: matches 3264..4457 of consensus"
repeat_region 29118..29521
/note="MIR repeat: matches 1..383 of consensus"
repeat_region 29523..30381
/note="LIMB8 repeat: matches 2496..3321 of consensus"
repeat_region 30382..30685
/note="AluJb repeat: matches 2..299 of consensus"
repeat_region 30686..31069
/note="LIMB8 repeat: matches 2110..2496 of consensus"
repeat_region 31088..31421
/note="LIMB8 repeat: matches 5820..6156 of consensus"
repeat_region 31422..31617
/note="LIMB8 repeat: matches 5374..5552 of consensus"
repeat_region 31618..31909
/note="AluJx repeat: matches 1..294 of consensus"
repeat_region 31910..31976
/note="LIMB8 repeat: matches 5304..5374 of consensus"
repeat_region 31978..32359
/note="LIMB8 repeat: matches 1951..2330 of consensus"
repeat_region 32354..33592
/note="LIMB8 repeat: matches 969..2197 of consensus"
repeat_region 33935..34790
/note="LIMB8 repeat: matches -266..602 of consensus"
repeat_region 34856..35219
/note="LIMB8 repeat: matches 1..364 of consensus"
repeat_region 36043..36159
/note="L2 repeat: matches 2576..2710 of consensus"
misc_feature 38455..39057
/note="match: STS: Em:G008362"
misc_feature 41978..42373
/note="match: STS: Em:G008362"
repeat_region 42056..42087
/note="L2 repeat: matches 2687..2742 of consensus"
repeat_region 42311..42540
/note="L2 repeat: matches 19..262 of consensus"
repeat_region 42824..42914
/note="L2 repeat: matches 2606..2709 of consensus"
repeat_region 43119..43173
/note="L2 repeat: matches 2687..2742 of consensus"
repeat_region 43122..43185
/note="MIR repeat: matches 194..257 of consensus"
repeat_region 44700..44743
/note="L2 repeat: matches 2687..2742 of consensus"
repeat_region 45956..46137
/note="MIR repeat: matches 20..199 of consensus"
repeat_region 46401..46617

```

```

repeat_region 46729..47035
/note="AluJb repeat: matches 1..309 of consensus"
repeat_region 47439..47682
/note="LIMB1 repeat: matches 5666..5914 of consensus"
repeat_region 48430..48527
Query Match 7.6%; Score 43.6; DB 9; Length 173510;
Best Local Similarity 67.8%; Pred. No. 0.066;
Matches 61; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 94 ctgagagagagagctccctgggtttactcctaaggtgtcctgagagagacga 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84025 CTGCAAGGGGGGCGACCTGGGCTTGAAGGGGGGCTGGAACACTGAGCCG 83966

OY 154 ttaatcatttaaggttcgaagaaggagc 183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83965 CTCACAGTGTCTAAGTAAGTAAGTGGCGC 83936

RESULT 15
LOCUS S39392 2111 bp mRNA PRI 05-MAR-2001
DEFINITION Homo sapiens protein tyrosine phosphatase mRNA, partial cds.
ACCESSION S39392
VERSION S39392.1 GI:250890
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2111)
AUTHORS Arikura, Y., Hinoda, Y., Itoh, F., Takekawa, M., Tsujisaki, M.,
Adachi, M., Imai, K. and Yachi, A.
CDNA cloning of new protein tyrosine phosphatases in the human
colon
JOURNAL Tumour Biol. 13 (3), 180-186 (1992)
MEDLINE 92327504
JOURNAL GENBANK staff at the National Library of Medicine created this
entry [NCBI gi250890] from the original journal article.
REMARK This sequence comes from Fig. 3.
FEATURES
source Location/Qualifiers
1..2111
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL6"
/issue_type="Colon"
<1..>2111
/note="pipase; This sequence comes from Fig. 3"
/codon_start=3
/product="protein tyrosine phosphatase"
/protein_id="AAB22439.2"
/db_xref="GI:13195754"
/translation="OHSGKQSEAESEYINARTLDFYGVELHSGRDLHNLDMTIGIA
SAGVAVYRKIKCSFPMYNNITKISPRKKKFFIHORQAESEHIVAFNMLNYSCK
NMRKSCVEHHTFPOAKLLPQENKNTSVNNQYKAGVGMVNPDA
NMRSISVEHLEKSLPSRSPITPMRSTRKHEIRKPHSSADNANAMTYTERED
VFETVYKSLAPQDSDEVSQNRSPHQSSENNPASQSYLQSSSSVSQNPAGSCS
PDGVDOOLLDDFHRVTRGSGTSDASQYCDKNDNGSYLVLRITPDEGKGFNKG
GVDOOKMPLVIRRAVRSPADKSPKLNESDQVILNGRDISHTHGVVVF1KASIES
HSRELAIVIRRAVRSPADKSPKLNESDQVILNGRDISHTHGVVVF1KASIES
GTVLIDPEQLYRKRPGLATITFATLPNDKNRYKQVLPDTRVLLGNGEDYINASYV
NMEIPANLVNKYIATQGLPHTCACQFQVYVMDQKSLVMTLTITLREGRTCHOYWP
DPPDVNMGHGFHIOCOSEDTIAYVSREMLVNTOTGEEHTVTHLOVYAMPDGHGTFDD
SDPLEFVNVRSLRVDSPEVLVHCSAGIGRTGVLTMTACTLTERNLPTPLDLYR
KMDORAMVQTSQYKVECE"
BASE COUNT 594 a 528 c 510 g 479 t
ORIGIN
Query Match 6.6%; Score 37.8; DB 9; Length 2111;
Best Local Similarity 53.8%; Pred. No. 2.9;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```

QY 97 gagggagagctccctgggtttttactcttaaaaggggtggcctggagcacgagagaccattta 156
 Db 969 GATCAAGATGAAATTTGGATTAACTTTAAGGGAGGAGTGGATCAAAAAGATGCCTCTT 1028
 QY 157 atcatctctaaggtcgaagaaggggcacaagcagacacacctgagctccaaactgcaggct 216
 Db 1029 GTGCTATCAAGGATTAACCCAGAGTCACCTGCGGACACCTGTCATTCTTAAGCTGAACGAA 1088
 QY 217 ggggatgaggttgcacatcaatg 241
 Db 1089 GGGATCAAAATCGTTAATCAATG 1113

Search completed: January 31, 2002, 19:45:37
 Job time: 1755 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:50:27 ; Search time 163.78 Seconds
(Without alignments)
2999.434 Million cell updates/sec

Title: US-09-641-831-1
Perfect score: 573
Sequence: 1 atgatgagaccacactgaaga.....gagccatttaagtagatga 573

Scoring table: IDENTITY_MNC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database : N.Geneseq_1101:*

1:	/SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	573	100.0	573 22 AAS00039	Human cDNA encodin
2	323	56.4	327 22 AAS00040	Human cDNA encodin
3	323	56.4	402 22 AAS00041	Human cDNA encodin
4	64.6	11.3	462 22 AAI37914	Probe #19760 used
5	55.8	9.7	114 22 AAI51074	Human pancreatic c
6	44.2	7.7	1525 21 AAC98998	Methylphilus meth
7	37.8	6.6	2366 21 AAC58627	Protein tyrosine P
8	37.8	6.6	3984 18 AAT58627	Human PRPH1 cDNA
9	37.4	6.5	717 20 AAT78463	Human bone mineral
	37.4	6.5	717 20 AAT24320	Human LIM mineral
			717 21 AAC87713	Human LIM mineral

12	37.4	6.5	1456 21 AAC87741	Human LMP-2 (HLM-
13	37.4	6.5	1575 21 AAC87742	Human LMP-3 (HLM-
14	37.4	6.5	1620 21 AAC87737	Human LMP express1
15	37.4	6.5	1644 20 AAX24322	Human bone mineral
16	37.4	6.5	1644 21 AAC87715	Human LIM mineral1
17	37.4	6.5	1665 20 AAX24328	Human truncated bo
18	37.4	6.5	1665 21 AAC87738	Human LIM 5'-UTR 5
19	37.4	6.5	1689 21 AAX24327	Human bone mineral
20	37.4	6.5	1689 21 AAC87727	Human LIM mineral1
21	36.6	6.4	1689 21 AAC75454	Human secreted pro
22	36.6	6.4	1689 21 AAC75454	Human secreted pro
23	36.6	6.4	1689 21 AAC75454	Human secreted pro
24	36.2	6.3	1696 21 AAC87708	Human secreted pro
25	35.8	6.2	376 21 AAC02040	Human secreted pro
26	34.8	6.1	325 20 AAX24324	Human bone mineral
27	34.2	6.0	765 20 AAX24324	Human bone mineral
28	34.2	6.0	765 21 AAC87726	Human secreted pro
29	34	5.9	1511 19 AAC73198	Human secreted pro
30	33.8	5.9	1593 19 AAC73198	Human secreted pro
31	33.8	5.9	2400 19 AAV19002	Human osteosarcoma
32	33.8	5.9	4768 21 AAT29097	Human normal blad
33	33.4	5.8	1597 17 AAT12908	Klebsiella aerogen
34	33.4	5.8	2162 20 AAX40058	Klebsiella aerogen
35	33.4	5.8	2236 20 AAX39644	Human insulin rece
36	33.4	5.8	2236 20 AAX40057	Human insulin rece
37	33.4	5.8	2289 20 AAX40061	Colon cancer assoc
38	33.4	5.8	2409 20 AAX40062	Colon cancer assoc
39	33.2	5.8	1336 19 AAV30952	Renal cancer assoc
40	33.2	5.8	1554 21 AAC75976	Human prostate can
41	33.2	5.8	1606 21 AAT16133	Human ORFX ORP131
42	33.2	5.8	1844 22 AAH18032	Human cDNA sequenc
43	33.2	5.8	2519 22 AAH16071	Human cDNA sequenc
44	33.2	5.8	2581 22 AAH16008	Human cDNA sequenc
45	33.2	5.8	2987 9 AAN80456	Sequence of Rhodosp

ALIGNMENTS

RESULT 1				
AA500039	standard; cDNA: 573 BP.			
ID	AA500039			
AC	AA500039			
XX				
DT	11-MAY-2001 (first entry)			
DE				
XX	Human cDNA encoding novel human protein, NHP#1.			
XX				
KW	Human; novel human protein; NHP#1; gene therapy; drug screening;			
KW	obesity; high blood pressure; ss.			
XX				
OS	Homo sapiens.			
XX				
FT				
FT				
FT				
XX				
XX				
PN	WC200114422-A1.			
PD	01-MAR-2001.			
XX				
PF	18-AUG-2000; 2000WO-US22815.			
XX				
PR	24-AUG-1999; 99US-0150511. <i>pre</i>			
XX				
PA	(LEXI-) LEXICON GENETICS INC.			
XX				
PI	Turner CA, Zambrowicz B, Friedlich G, Nehls M, Sands AT;			
XX				
DR	WPI: 2001-218430/22.			
DR	P-PSDB: AAU00031.			

XX	11-MAY-2001	(first entry)
XX		
XX	Human cDNA encoding novel human protein, NHP#2.	
XX		
XX	Human; novel human protein; NHP#2; gene therapy; drug screening;	
KW	obesity; high blood pressure; ss.	
KW		
XX		
XX	Homo sapiens.	
OS		
XX		
FH	Key	Location/Qualifiers
FF	CDS	1..327
FT		/tag= a
FT		/product= "NHP#2"
FT		
XX		
PN	WO200114422-A1.	
XX		
PD	01-MAR-2001.	
XX		
PF	18-AUG-2000; 2000WO-US22815.	
XX		

2X
PA
XX
PI
PI
PI

(LEXI-) LEXICON GENETICS INC.
Turner CA, zambrowicz B, Friedrich G, Nehls M, Sands AT;

WPI: 2001-218430/22.
P-PSDB; AAU00032.

Novel human polynucleotides isolated from human mammary gland cDNA library, encodes novel human proteins which are useful in diagnosis, drug screening, clinical trial monitoring or treating behavioural disorders

Claim 3; Page 25; 29pp; English.

The sequence encodes a novel human protein, NHP#2, which shares structural motifs with human APXL protein. Nucleotide constructs encoding functional NHPs are used in gene therapy approaches for the modulation of NHP expression. NHP oligonucleotides can be used as hybridisation probes for screening libraries and assessing NHP gene expression patterns. Also, labeled NHP nucleotide probes can be used to screen a human genomic library. The NHP nucleotide sequences are also useful in drug screening techniques for treating symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. Examples of such manifestations may include obesity and high blood pressure.

Sequence 327 BP; 91 A; 76 C; 96 G; 64 T; 0 other;

Query Match 56.4%; Score 323; DB 22; Length 327;
Best Local Similarity 100.0%; Pred. No. 9.5e-93;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 atgatggagccactgaagacttcacaaagcctagtgcacattaaactaacacgccc 60
| | | | |
Db 1 atgatggagccactgaagacttcacaaagcctagtgcacattaaactaacacgccc 60
| | | | |
Qy 61 accaagggaaggtcacattatctggagcattctctggaggagagagctccctggggtttt 120
| | | | |
Db 61 accaagggaaggtcacattatctggagcattctctggaggagagagctccctggggtttt 120
| | | | |
Qy 121 actctaaaggtggcctgggacgaggaaccattaatcctctaaaggtcgagaaggg 180
| | | | |
Db 121 actctaaaggtggcctgggacgaggaaccattaatcctctaaaggtcgagaaggg 180
| | | | |
Qy 181 ggcaaaagcagacacccctgagctccaaactcgagctggggatgaggttctgcacataaat 240
| | | | |
Db 181 ggcaaaagcagacacccctgagctccaaactcgagctggggatgaggttctgcacataaat 240
| | | | |
Qy 241 gagtgactctgacacactccagaagaagagcaatttcccttggtgaaagggatcctacaag 300
| | | | |

181 ggcaagcagacacccctgagctccaaactcgagctggggatgaggttgtgcacatcaat 240
 |||||
 181 ggcaagcagacacccctgagctccaaactcgagctggggatgaggttgtgcacatcaat 240
 |||||
 241 gaggtagctcgacagctccagaaagagacagtttcccttgtgaaaggaatcctacaag 300

Db 241 gaggtgactctgagcagctcccaagaaggagcagcttcctctgtgaaagatcctacaag 300
 QY 301 accctcagcgtgtagtgagcag 323
 Db 301 accctcagcgtgtagtgagcag 323

RESULT 3

AS00041 ID AS00041 standard; cDNA; 402 BP.

AC AAS00041;

DE 11-MAY-2001 (first entry)

XX Human cDNA encoding novel human protein, NHP#3.

XX Human; novel human protein; NHP#3; gene therapy; drug screening;

KW Obesity; high blood pressure; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1.402

XX MO200114422-A1.

XX 01-MAR-2001.

XX 18-AUG-2000; 2000MO-US22815.

XX 24-AUG-1999; 99US-0150511.

XX (LEXI-) LEXICON GENETICS INC.

XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;

XX WPI; 2001-218430/22.

XX P-SDB; AAU00033.

XX Novel human polynucleotides isolated from human mammary gland cDNA

XX library, encodes novel human proteins which are useful in diagnosis,

XX drug screening, clinical trial monitoring or treating behavioural

XX disorders

XX Claim 5; Page 25; 29pp; English.

XX The sequence encodes a novel human protein, NHP#3, which shares

XX structural motifs with human APX1 protein. Nucleotide constructs

XX encoding functional NHPs are used in gene therapy approaches for the

XX modulation of NHP expression. NHP oligonucleotides can be used as

XX hybridisation probes for screening libraries and assessing NHP gene

XX expression patterns. Also, labeled NHP nucleotide probes are also

XX useful in drug screening techniques for treating symptomatic or

XX phenotypic manifestations of perturbing the normal function of NHP in the

XX body. Examples of such manifestations may include obesity and high blood

XX pressure.

XX Sequence 402 BP; 110 A; 91 C; 117 G; 84 T; 0 other;

Query Match 56.4%; Score 323; DB 22; Length 402;
 Best Local Similarity 100.0%; Pred. No. 1e-92;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 atgatgagacacactgaagactccacaagcctagtcgacatttaactctaacgcgc 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 atgatgagacacactgaagactccacaagcctagtcgacatttaactctaacgcgc 60

QY 61 accaagggaaggtacatttattctgtgagcattcctctgaggaaggagctccctggggttt 120
 Db 61 accaagggaaggtacatttattctgtgagcattcctctgaggaaggagctccctggggttt 120
 QY 121 actctaaagggtgctctgagcagcaggaacattatctcttaaggctcgaagaagg 180
 Db 121 actctaaagggtgctctgagcagcaggaacattatctcttaaggctcgaagaagg 180
 QY 181 ggcgaagcagacacacctgagctcccaactcagcgtggtggtatgtttgacatcaat 240
 Db 181 ggcgaagcagacacacctgagctcccaactcagcgtggtggtatgtttgacatcaat 240
 QY 241 gaggtgactctgagcagctccagaagaaggagcagcttcctctgtgaaaggatcctaag 300
 Db 241 gaggtgactctgagcagctccagaagaaggagcagcttcctctgtgaaaggatcctaag 300
 QY 301 accctcagcgtgtagtgagcag 323
 Db 301 accctcagcgtgtagtgagcag 323

RESULT 4

AI37914 ID AI37914 standard; DNA; 462 BP.

AC AI37914;

DE 17-OCT-2001 (first entry)

XX Probe #6600 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

XX Homo sapiens.

XX MO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta

XX Claim 25; SEQ ID No 6600; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).

XX producing a microarray for predicting, measuring and displaying gene

XX expression in samples derived from human placenta. The probes are useful

XX for antenatal diagnosis of human genetic disorders.

Query Match 11.3%; Score 64.6; DB 22; Length 462;
 Best Local Similarity 65.7%; Pred. No. 1.3e-10;
 Matches 94; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

1 atgatgagacacactgaagactccacaagcctagtcgacatttaactctaacgcgc 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 atgatgagacacactgaagactccacaagcctagtcgacatttaactctaacgcgc 60

Db	62	gcgatttcctggtgaagggtcccataagaccctgaagctggtcgt	108
RESULT	6		
	AAC98998		
	ID	AAC98998 standard; cDNA; 1525 BP.	
	XX		
	XX		
	AC		
	AC		
	XX		
	XX		
	DT	09-MAR-2001 (first entry)	
	XX		
	XX		
	DE	Human pancreatic cancer antigen nucleotide sequence SEQ	

Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection; diagnosis; identification; cytosolic; neuroprotective; neurotropic; immunomodulatory; relaxant; contraceptive; gynaecological; antiinflammatory; cardiac; gene therapy; chromosome mapping; linkage analysis; tissue identification; tissue typing; forensic; neural; immune system; muscular; reproductive; gastrointestinal; pulmonary; cardiovascular; renal; proliferative; ss.

OS Homo sapiens.

XX
PN WO200055320-A1.

XX PD 21-SEP-2000.

XX
PF
08-MAR-2000:XX
PR 12-MAR-1999:XX
PA (HITMA -) HITMAN G

XX
PT
Rosen CA
Rube

LI ROSEN CA, KADE
XX
DB FBI: 3000-57044

DR
DR
VY

PT New nucleic acid

PT treating, or am
PT cancer, or for

XX PS Claim 1; Page 6

XX
CC AAC98773 to AAC

CC proteins, called
CC AAB54466. The h

CC neuroprotective
CC gynaecological,

CC in gene therapy
CC preventing, tre

CC for diagnosing
CC subject. Binding

CC proteins can be
CC detect, treat o

CC Agonists and antipneumonic cancer

acid hybridisation analysis, tissue

and diagnostic methods, which are used to

both in vivo and

proteins can be
reproductive, ga
analyses

CC
CC
CC

XX
SQ
Sequence 1525 B1

Query Match
Best Local Similarity

Matches 73; **Cons**

100

Query Match	6.68;	Score 37.8;	DB 21;	Length 2566;
Best Local Similarity	58.48;	Pred. No. 0.096;		
Matches 66;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0

Db 1142 accgcacccagcaggtgctgcgcgcacgcttgcgcaaaaagcgcctatttcctg 1194

RESULT	8
AAT58627	
ID	AAT58627 standard; cDNA; 3984 BP

AC AAT58627;

DT 22-MAY-1997 (first entry)

DE Protein tyrosine phosphatase cDNA

KW protein tyrosine phosphatase; PTPN1; focal adhesion

retrovirus; vector; ss.

OS Homo sapiens

Key	Location/Qualifiers
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

33

US5595911-A.

AA PD 21-JAN-1997

XX
PF 14-MAR-1990

01-MAR-1991
PR
AA

PR 16-AUG-1993

XX PA (COTD-) COTD

Tonks NK:
XX
PI

XX
DE
EPT: 1997-11

DR P-PSDB; AAW.

PT DNA encoding

XX
2
1

XX

CC (PTPH1) (AA

action of a

CC	CDNA clone
CC	10001-10002

CC conserved
CC a PCR produ

CC de incuþiei
CC de expresiei

CC overcome or
CC residues at

CC the prevent
XX

Sequence 39

Query Match	6.6%;	Score 37.8;	DB 18;	Length 3984;
Best Local Similarity	53.8%;	Pred. No. 0.12;		
Matches	78;	Conservative	0;	Mismatches 67; Indels 0; Gaps 0;

QY	97	gaggagagagctccctgggggttttactctaaagggtggcctggagcagcgagaaccatta	156
Db	1569	gatgaagatggaaaattggatttaatttaagggaggagtgatgataaagatgcctctt	1628
QY	157	atcatctctaaagtgcgaagaagggggcaaacagacacacctgagctccaaactcaggct	216
Db	1629	gtggtatcaagatgaataaacccagagtcacctggaacacctgattctctaagctgaacgaa	1688
QY	217	ggggatggaggtgtgtgcacatcaatg	241
Db	1689	ggggatcaaatcggtttaatcaatg	1713

RESULT	9
AAX78463	
ID	AAX78463 standard; cDNA; 3984 BP.
XX	AC
XX	AAX78463;
XX	07-SEP-1999 (first entry)
XX	DT
XX	DE
XX	Human PTPH1 cDNA.
KW	PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;
KW	localisation; treatment; overexpression; oncogenic; cell transformation;
KW	prevention; phosphotyrosine; disease; malignant; ss.
XX	OS
XX	Homo sapiens.
XX	Key
XX	CDS
FT	Location/Qualifiers
FT	24..2765
FT	/*tag= a
FT	/product= "PTPH1"
FT	111..1094
FT	/*tag= b
FT	/note= "Region of homology to the N-terminal domain
FT	of band 4.1, ezrin and talin. This region is
FT	known to be important for localisation to focal
XX	adhesions"
PN	US5863781-A.
XX	XX
XX	26-JAN-1999.
XX	XX
XX	04-DEC-1996; 96US-0759536.
XX	XX
PR	01-MAR-1991; 91US-0663579.
PR	14-MAR-1990; 90US-0494036.
PR	16-AUG-1993; 93US-0107420.
PR	04-DEC-1996; 96US-0759536.
XX	XX
XX	(COLD-) COLD SPRING HARBOR LAB.
XX	PA
XX	XX
PI	Tonks NK;
XX	XX
XX	WPI; 1999-131308/11.
DR	P-PSDB; AAY25156.
DR	XX
PT	Protein tyrosine phosphatase PTPH1 - encoded by DNA of HeLa cells
PT	XX
PS	Disclosure; Fig 1A-B; 12pp; English.
PS	XX
CC	CC
CC	This sequence encodes a novel protein tyrosine phosphatase, PTPH1
CC	isolated from HeLa cells. The protein of the invention appears to
CC	localise to focal adhesions and is therefore potentially useful in the
CC	treatment of cancer. Overexpression of PTPH1 can be used to counter the
CC	effects of oncogenic protein tyrosine kinases such as those of
CC	transforming viruses and for interfering with or reversing cell

CC easier assessment of transfected cells. Clinical applications
 CC include enhancement of bone repair in fractures, bone defects, bone
 CC grafting, and normal homeostasis in patients presenting with
 CC osteoporosis. A method of inducing bone formation using
 CC transfected osteogenic precursor cells is claimed.

XX Sequence 717 BP; 138 A; 262 C; 219 G; 98 T; 0 other;

Query Match

Best Local Similarity 49.8%; Score 37.4; DB 20; Length 717;
 Matches 126; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

QY 86 aggcattcttgaggagagagctccctgggttttacttaaaagggtgctggagcagc 145
 DB 14 agtgatgctggaggagagagcagcacttggtctccggtgcaaggagagacttca 73
 QY 146 gagaacattatcatctcctaagtcgaaggaggggcaaacagacaccctgagctcca 205
 DB 74 atgtgccctctccattctcccgctcactcctctgggggcaaaagcgc-----gcaggccg 127
 QY 206 aactgcagctggaggatgaggtgtgcacatcaatgaggtgactctgagcagctccagaa 265
 DB 128 gaggcgccgtgggtgactgaggtgtgcagatcgatgagcagaatgctggtagcctcacac 187
 QY 266 aggaagcgagtttccctgggtgaagagatcctacaagaccctcagagctgtgtagcgagcc 325
 DB 188 acatcgagagctcagaacaagatccgcgctcgaggagcgctcagctggtgctcagca 247
 QY 326 tctcccaaccggt 338
 DB 248 gggcccaagcgggt 260

RESULT 11

ID AAC87713 standard; cDNA; 717 BP.

AC AAC87713;

XX 01-MAR-2001 (first entry)

DE Human LIM mineralisation protein (HLMF) partial cDNA seq ID NO:7.

XX LIM mineralisation protein; LMP; bone formation; osteopathic;

KW osteogenic precursor cell; gene therapy; metabolic bone disease;

KW osteoporosis; bone degenerative disease; ss.

OS Homo sapiens.

XX WO200066178-A1.

PD 09-NOV-2000.

XX 28-APR-2000; 2000MO-US11664.

XX 30-APR-1999; 99US-0132021.

XX (UYEM-) UNIV EMORY.

XX Boden SD, Hair GA;

XX WPI; 2000-672828/65.

XX New nucleic acid encoding a human LIM mineralization protein for
 PT inducing or inhibiting bone formation, fusing a spine, stimulating
 PT production of an osteogenic cell, or treating bone conditions, such as,
 PT osteoporosis

XX Example 17; Page 61; 84pp; English.

XX present invention specifically describes the human LIM
 XX mineralisation proteins (LMP) HLMF-2 and HLMF-3. LMPs have osteopathic

CC activity and can be used in gene therapy. LMP nucleic acids can be used
 CC to induce or inhibit bone formation, fuse a spine, stimulate production
 CC of an osteogenic cell, or inhibit the expression of HLMF-2 or HLMF-3.
 CC They can be used to treat bone conditions, such as, osteoporosis and
 CC other metabolic bone diseases. Antibodies to the LMP proteins encoded
 CC by the nucleic acids are used in marker assays to identify risk factors
 CC in bone degenerative diseases, such as osteoporosis. The nucleic acids
 CC are used in gene therapy for bone formation which leads to the
 CC advantages of: (1) lower production costs; (2) greater efficacy compared
 CC to extracellular treatment regimens due to the ability to achieve
 CC prolonged expression of the intracellular signal; (3) by-passing the
 CC possibility that treatment with extracellular signals might be hampered
 CC due to the presence of limiting numbers or receptors for those signals;
 CC (4) permitting the delivery of transfected potential osteoprogenitor
 CC cells directly to the site where localised bone formation is required;
 CC and (5) permitting systemic bone formation, which provides a treatment
 CC regimen for osteoporosis and other metabolic bone diseases. The present
 CC sequence represents a partial human LMP cDNA sequence, which is given
 CC in an example from the present invention.

XX Sequence 717 BP; 138 A; 262 C; 219 G; 98 T; 0 other;

Query Match

Best Local Similarity 49.8%; Score 37.4; DB 21; Length 717;
 Matches 126; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

QY 86 aggcattcttgaggagagagctccctgggttttacttaaaagggtgctggagcagc 145
 DB 14 agtgatgctggaggagagagcagcacttggtctccggtgcaaggagagacttca 73
 QY 146 gagaacattatcatctcctaagtcgaaggaggggcaaacagacaccctgagctcca 205
 DB 74 atgtgccctctccattctcccgctcactcctctgggggcaaaagcgc-----gcaggccg 127
 QY 206 aactgcagctggaggatgaggtgtgcacatcaatgaggtgactctgagcagctccagaa 265
 DB 128 gaggcgccgtgggtgactgaggtgtgcagatcgatgagcagaatgctggtagcctcacac 187
 QY 266 aggaagcgagtttccctgggtgaagagatcctacaagaccctcagagctgtgtagcgagcc 325
 DB 188 acatcgagagctcagaacaagatccgcgctcgaggagcgctcagctggtgctcagca 247
 QY 326 tctcccaaccggt 338
 DB 248 gggcccaagcgggt 260

RESULT 12

ID AAC87741 standard; cDNA; 1456 BP.

AC AAC87741;

XX 01-MAR-2001 (first entry)

DE Human LMP-2 (HLMF-2) encoding cDNA sequence SEQ ID NO:37.

XX LIM mineralisation protein; LMP; bone formation; osteopathic;

KW osteogenic precursor cell; gene therapy; metabolic bone disease;

KW osteoporosis; bone degenerative disease; ss.

OS Homo sapiens.

XX WO200066178-A1.

XX 09-NOV-2000.

XX 28-APR-2000; 2000MO-US11664.

XX 30-APR-1999; 99US-0132021.

XX (UYEM-) UNIV EMORY.

LIM mineralisation protein; LMP; bone formation; osteopathic;
osteogenic precursor cell; gene therapy; metabolic bone disease;
osteoporosis; bone degenerative disease; ss.

Homo sapiens.
WO200066178-A1.
09-NOV-2000.
28-APR-2000; 2000WO-US11664.
30-APR-1999; 99US-0132021.
(UYEM-) UNIV EMORY.
Boden SD, Hair GA;
WPI; 2000-672828/65.
P-PSDB; AAB36475.

New nucleic acid encoding a human LIM mineralization protein for
inducing or inhibiting bone formation, fusing a spine, stimulating
production of an osteogenic cell, or treating bone conditions, such as,
osteoporosis

Claim 1; Page 44-45; 84pp; English.

The present invention specifically describes the human LIM
mineralisation proteins (LMP) HLMF-2 and HLMF-3. LMPs have osteopathic
activity and can be used in gene therapy. LMP nucleic acids can be used
to induce or inhibit bone formation, fuse a spine, stimulate production
of an osteogenic cell, or inhibit the expression of HLMF-2 or HLMF-3.
They can be used to treat bone conditions, such as, osteoporosis and
other metabolic bone diseases. Antibodies to the LMP proteins encoded
by the nucleic acids are used in marker assays to identify risk factors
in bone degenerative diseases, such as osteoporosis. The nucleic acids
are used in gene therapy for bone formation which leads to the
advantages of: (1) lower production costs; (2) greater efficacy compared
to extracellular treatment regimens due to the ability to achieve
prolonged expression of the intracellular signal; (3) by-passing the
possibility that treatment with extracellular signals might be hampered
due to the presence of limiting numbers or receptors for those signals;
(4) permitting the delivery of transfected potential osteoprogenitor
cells directly to the site where localised bone formation is required;
and (5) permitting systemic bone formation, which provides a treatment
regimen for osteoporosis and other metabolic bone diseases. The present
sequence encodes the human LMP-2 (HLMF-2) amino acid sequence from
the present invention.

Sequence 1456 BP; 288 A; 469 C; 459 G; 240 T; 0 other;

Query Match 6.5%; Score 37.4; DB 21; Length 1456;
Best Local Similarity 49.8%; Pred. No. 0.099;
Matches 126; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

QY 86 aggcattcctggaggagagtcctccgtggttttactctaagggtgctggagcacg 145
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 agtgtagctggaggccacaccttgggttcggtcgaaaggggcaaggacttca 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 146 gagaaccattaatcatcttaagtgaaggaagggggcaaacagacacctgagctcca 205
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 atgtgccctctccattcccgcctcactcctctgggggcaagcgcc-----gcaggccg 172
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 206 aactgcaggctgggatgaggttgtgcatacaatgagtgactctgagcagctccagaa 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 gagtggcctgggtgactggtgctgagcatcgatggcgaaatgcgggtgagctccac 232
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 266 aggagcaggttccctggtgaagagatcctcaagaccctcagctggttagtgcgagcc 325
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 acatcgaagctcagaacaagatccggcctgctggggagcgctcagctgggctcagca 292
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 ggccacagcgggt 305
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AAC87742
ID AAC87742 standard; cDNA; 1575 BP.
XX AAC87742;
AC AAC87742;
AC AAC87742;
XX 01-MAR-2001 (first entry)
DT Human LMP-3 (HLMF-3) encoding cDNA sequence SEQ ID NO:39.
DE DE
XX XX

Qy	326	tctcccaaccgt	338
Db	293	gggcccaagccgt	305

RESULT 14
A0007727

ID AAC87737 standard; cDNA; 1620 BP.

AC AAC87737

DT 01-MAR-2001 (first entry)

Accession	Human LMP expression vector cDNA sequence SEQ ID NO:32
DE	

AA LIM mineralisation protein; LMP; bone formation; osteopathic;
KM osteogenic precursor cell; gene therapy; metabolic bone disease;
KM osteoporosis; bone degenerative disease; ss.

OS Homo sapiens.

PN WO2000066178-A1

09-NOV-2000 PD

PF 28-APR-2000; 2000WO-US11664

PR 30-APR-1999; 99US-0132021

PA (UYEM-) UNIV EMORY

PI Boden SD, Hall GA.

DR WPI; 2000-672828/65.

aa New nucleic acid encoding a human LIM mineralization protein for
pt inducing or inhibiting bone formation, fusing a spine, stimulating
pt production of an osteogenic cell, or treating bone conditions, such as
pt osteoporosis -

PS Example 23; Page 69; 84pp; English.

The present invention specifically describes the human LIM mineralisation proteins (LMP) HLMF-2 and HLMF-3. LMP nucleic acids can be used to induce or inhibit bone formation, use a spine, stimulate production of an osteogenic cell, or inhibit the expression of HLMF-2 or HLMF-3. They can be used to treat bone conditions, such as, osteoporosis and other metabolic bone diseases. Antipodotes to the LMP proteins encoded by the nucleic acids are used in marker assays to identify risk factors in bone degenerative diseases, such as osteoporosis. The nucleic acids are used in gene therapy for bone formation which leads to the advantages of: (1) lower production costs; (2) greater efficacy compared to extracellular treatment regiments due to the ability to achieve prolonged expression of the intracellular signal, (3) by-passing the possibility that treatment with extracellular signals might be hampered due to the presence of limiting numbers or receptors for those signals; (4) permitting the delivery of transcribed potential osteoprogenitor cells directly to the site where bone formation is required; and (5) permitting systemic bone formation, which provides a treatment regimen for osteoporosis and other metabolic bone diseases. The present sequence represents a human LMP expression vector sequence, which is given in an example from the present invention.

sq sequence 1620 BP; 315 A; 558 C; 480 G; 267 T; 0 other;

Query Match	6.5%	Score 37.4	DB 21	Length 1620
Best Local Similarity	49.8%	Pred. No. 0.1		
Matches 166	Conservative	0	Mismatches 121	Indels 6
				Gaps 1

86 aggcattcctgagggagagagctccctctgggttctaactctaaaggctgcctgagacag 145

Db	14	aggtcagtgcctgtagggcgccaaacacttcgggcttccgcgcgcgaaggggcaagaagtcca	73
Qy	146	gagaaaccataatcatcattctaaggctcgaaagaagggcgcaaaagcacgaaccttgatccc	205
Dd	74	atgtgcgccctccatcattccgcgcgtccactccctgggggcaaaagcggc-----gaaagcgg	127
Qy	206	aactcagaagcgcggagtaggtctgtgcacatcaaatcgagcttgactctgagcagctccagaa	265
Dd	128	gagctvgccgctgvgcgaactcgggtgtctgcagacatcacgatgtgagaaatcgggtttacac	187
Qy	266	aggagagcagcttccctcgtctgaagaagatccctacaagaaccttaagctctgttagtgcagagc	324
Dd	188	acatcagaagcctaagaacaagaatccgcggcctctcggggagagcgtcctaagccttgacctca	24
Qy	326	tctccccaacgggt	338
Dd	248	gggcccaagcgggt	260

RESULT 15

ID AAX24322 standard; cDNA; 1644 BP

AC AAX24322

DT 07-JUN-1999 (first entry)

DE Human bone mineralisation protein LMP CDNA

KW LMP; LM mineralisation protein; bone mineralisation; human
 KW cytokine; spine fusion; fracture repair; bone grafting;
 KW osteoporosis; gene therapy; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
1	137A	

ET

ET.

PN WO9906563-A1

PD 11-FEB-1999.

29-JUL-1998; 98WMO-US15814

PR 02-APR-1998; 98US-008040/
PR 30-JUL-1997; 97US-0054219

XX
PA (TIVEM-) TINTV EMORY.

XX Boden S, Hair G;
PT

XX
DR WPT: 1999-153793

P-PDB, ADAM2/044.
DK
XX

PT osteogenic precu

pt
structures
xx
ps
Example 19; Page 50; 67pp; English.

AA This nucleotide sequence encodes a
CC

enhances bone mineralization in mammalian cells grown *in vitro*, and also induces bone formation *in vivo* when produced in mammals. The sequence was obtained from 2 overlapping cDNA clones (see AAX24330 and AAX24321) obtained from human osteosarcoma MG63 and human heart. A 5' untranslated region has also been obtained (see AAX24327), unlike other cytokines, LMP is not a secreted protein, but is instead an intracellular signalling molecule, so can provide intracellular signalling amplification and easier assessment of

